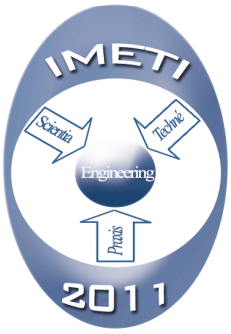



# The 4th International Multi-Conference on Engineering and Technological Innovation

July 19th - July 22nd, 2011 – Orlando, Florida, USA

# PROCEEDINGS

## Volume II

### (Post-Conference Edition)

**Edited by:**

**Nagib Callaos**
**José Vicente Carrasquero**
**Hsing-Wei Chu**
**José Ferrer**
**Michael J. Savoie**
**Andrés Tremante**


**Organized by**

**International Institute of Informatics and Systemics**

**Member of the International Federation for Systems Research (IFSR)**


## COPYRIGHT



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 no necessarily constitute endorsement by the editors.

ISBN- 978-1-936338-35-1 (Collection)

ISBN-978-1-936338-46-7 (Volume II (Post-Conference))


# PROGRAM COMMITTEE

**Chairs:** Hsing-Wei Chu (USA) and C. Dale Zinn (USA)

| | | |
|---|---|---|
| Adascalitei, Adrian | Technical University | Romania |
| Aguilar Torres, Fernando J. | Almeria University | Spain |
| Ahmad, Ashraf | Princess Sumya University for Technology | Jordan |
| Al Obaidy, Mohaned | Gulf College | Oman |
| Al-Aomar, Raid | Jordan University of Science and Technology | Jordan |
| Alshara, Osama | Higher Colleges of Technology | Uzbekistan |
| Ariwa, Ezendu | London Metropolitan University | UK |
| Audant, A. Babette | CUNY Kingsborough Community College | USA |
| Ayuga Téllez, Esperanza | Universidad Politécnica de Madrid | Spain |
| Barberà, Elena | Universitat Oberta de Catalunya | Spain |
| Batos, Vedran | University of Dubrovnik | Croatia |
| Benbouziane, Mohamed | University of Tlemcen | Algeria |
| Bernardino, Jorge | Polytechnic Institute of CoimbraISEC – Instituto Superior de Engenharia de Coimbra | Portugal |
| Bönke, Dietmar | Reutlingen University | Germany |
| Braiman, Avital | Brown University | USA |
| Bubnov, Alexey | Institute of Physics | Czech Republic |
| Calenzo, Patrick | IM2NP | France |
| Carbon, Margarita | The Boeing Company | USA |
| Castaño-Moraga, Carlos | Universidad de Las Palmas de Gran Canaria | Spain |
| Chang, Maiga | National Science and Technology Program for e-Learning | Taiwan |
| Chang, Ruay-Shiung | National Dong Hwa University | Tokelau |
| Chang, Wen-Kui | Tunghai University | Taiwan |
| Chauhan, K. A. | S.V.National Institute of Technology-Surat | India |
| Chen, Yuhua | University of Houston | USA |
| Chen, Zhi | University of Kentucky | USA |
| Cheng, Xiaochun | University of Reading | UK |
| Chinesta, Francisco | ENSAM | France |
| Chiou, Yin-Wah | Nanhua University | Taiwan |
| Clarke, Tim | University of Wales | UK |
| Cone, Chris | Georgia Institute of Technology | USA |
| Curran, Kevin | University of Ulster | UK |
| Davies, B. J. | IJAMT | UK |
| Davim, J. Paulo | University of Aveiro | Portugal |
| Dierneder, Stefan | Linz Center of Mechatronics Gmbh | Austria |
| Dodig-Crnkovic, Gordana | Mälardalen University | Sweden |

| | | |
|---|---|---|
| Dudek, Agnieszka | Wroclaw University of Environmental | Poland |
| Ehmann, Kornel F. | Northwestern University | USA |
| Erkollar, Alptekin | University of Applied Sciences Wiener Neustadt | Austria |
| Estrems, Manuel | Universidad Politécnica de Cartagena | Spain |
| Faura, Félix | Universidad Politécnica de Cartagena | Spain |
| Forbes, Alex | Georgia Institute of Technology | USA |
| Franco, Patricio | Universidad Politécnica de Cartagena | Spain |
| Fu, Yonggang | Shanghai Jiao Tong University | China |
| Fúster-Sabater, Amparo | CSIC | Spain |
| Gelbukh, Alexander | National Polytechnic Institute | Mexico |
| Gibson, Thomas L. | General Motors Research and Development Center | USA |
| Glovnea, Romeo P. | Brunel University | UK |
| Gomes, Samuel | UTBM | France |
| Goriachkin, Oleg | Volga State Academy of Telecommunication and Informatics | Russian Federation |
| Greenberg, William | Virginia Tech | USA |
| Grenmyr, Gustav | Chalmers University of Technology | Sweden |
| Gschweidl, Manfred | University of Applied Sciences Vorarlberg | Austria |
| Hepdogan, Seyhun | University of Central Florida | USA |
| Hetzer, Dirk | T-Systems Nova Deutsche Telekom | Germany |
| Higashiyama, Yoichi | Ehime University | Japan |
| Hirschberg, Wolfgang | Graz University of Technology | Austria |
| Holifield, David | University of Wales | UK |
| Huang, Yo-Ping | Tatung University | Taiwan |
| Ioannides, Stathis | SFK | Netherlands |
| Iovan, Stefan | Romanian Railway | Romania |
| Jiménez Rodríguez, Lourdes | University of Alcala | Spain |
| Jingchao, Chen | University Donghua | China |
| Kalogiannis, Konstantinos | Brunel University | UK |
| Kaminski, Jacek | Chalmers University of Technology | Sweden |
| Keiski, Riitta L. | University of Oulu | Finland |
| Khamba, J. S. | Punjabi University | India |
| Kobielarz, Magdalena | Wroclaw University of Technology | Poland |
| Koeglmayr, Hans-Georg | Pforzheim University of Applied Sciences | Germany |
| Koleva, Maria | Bulgarian Academy of Sciences | Bulgaria |
| Komerath, Narayanan | Georgia Institute of Technology | USA |
| Kuropka, Piotr | Wroclaw University of Technology | Poland |
| Lahlouhi, Ammar | University of Biskra | Algeria |
| Lamo, Yngve | Bergen University College | Norway |
| Lappas, Georgios | TEI of Western Macedonia | Greece |
| Li, Jackie | The City College of the City University of New York | USA |
| Li, Longzhuang | Texas A & M University | USA |
| Li, Man-Sze | IC Focus Ltd. | UK |
| Litvin, Vladimir | California Institute of Technology | USA |
| Liu, Jun | University of Ulster | UK |
| Lloret Mauri, Jaime | Polytechnic University of Valencia | Spain |
| López Román, Leobardo | University of Sonora | Mexico |
| Luh, Guan-Chun | Tatung University | Taiwan |
| Macke, Janaina | Universidade de Caxias do Sul | Brazil |
| Mansikkamäki, Pauliina | Tampere University of Technology | Finland |
| Mares, Cristinel | Brunel University | UK |

| | | |
|---|---|---|
| Masoumi, Nasser | University of Tehran | Iran |
| Masunov, Artëm E. | University of Central Florida | USA |
| Mbobi, Aime Mokhoo | Red Knee Inc. | Canada |
| Mehrabian, Ali | University of Central Florida | USA |
| Mikhaylov, Ivan A. | University of Central Florida | USA |
| Mosley, Pauline | Pace University | USA |
| Mueller, Christoph Jo. | INCAP GmbH | Germany |
| Mura, Gianluca | Politecnico di Milano University | Italy |
| Murugan, Natarajan | Coimbatore Institute of Technology | India |
| Nagai, Yasuo | Tokyo University of Information Sciences | Japan |
| Nagar, Atulya K. | Liverpool Hope University | UK |
| Narasimhan, V. Lakshmi | Western Kentucky University | USA |
| Newman, Stephen | University of Bath | UK |
| Ngabonziza, Yves | LAGCC of the City University of New York | USA |
| Nicolescu, Cornel Mihai | The Royal Institute of Technology | Sweden |
| Nippa, Markus | Pforzheim University of Applied Sciences | Germany |
| Oberer, Birgit | University of Klagenfurt | Austria |
| Ong, Pangleen | University of Kentucky | USA |
| Orosco, Henry | University of Houston | USA |
| Ouwerkerk, David B. | General Motors Advanced Technology Center | USA |
| Paredes, Julia | GGC | USA |
| Patel, Anant | Georgia Institute of Technology | USA |
| Pennington, Richard | Georgia Gwinnett College | USA |
| Pérez, Carlos A. | Colombian Institute of Petroleum | Colombia |
| Petit, Frédéric | École Polytechnique de Montréal | Canada |
| Podaru, Vasile | Academia Tehnica Militara | Romania |
| Potorac, Alin Dan | University of Suceava | Romania |
| Praus, Petr | Charles University | Czech Republic |
| Pursell, David | Georgia Gwinnett College | USA |
| Rai, Bharatendra K. | University of Massachusetts | USA |
| Ramessur, Roshan | University of Mauritius | Mauritius |
| Remondino, Marco | University of Turin | Italy |
| Revetria, Roberto | Università Degli Studi di Genova | Italy |
| Rieder, Mathias | University of Applied Sciences Vorarlberg | Austria |
| Rodger, P. M. | University of Warwick | UK |
| Rodríguez Lozano, Gloria I. | National University of Colombia | Colombia |
| Rossbacher, Patrick | Graz University of Technology | Austria |
| Rudakov, Fedor | Oak Ridge National Laboratory | USA |
| Sahli, Nabil | Ministry of Technologies of Communication | Tunisia |
| Sala, Nicoletta | Universitá della Svizzera Italiana | Italy |
| Samtani, B. K. | S.V.National Institute of Technology-Surat | India |
| Sarate, João Alberto Rubim | Centro de Ensino Superior Cenecista de Farroupilha | Brazil |
| Sauder, Deborah | Georgia Gwinnett College | USA |
| Schaeffer, Donna M. | Marymount University | USA |
| Schaetter, Alfred | Pforzheim University | Germany |
| Schulman, Stuart | CUNY Kingsborough Community College | USA |
| Schumacher, Jens | University for Applied Sciences Vorarlberg | Austria |
| Shiraishi, Yoshiaki | Nagoya Institute of Technology | Japan |
| Siddique, Mohammad | Fayetteville State University | USA |
| Singh, Harwinder | Guru Nanak Dev Engineering College | India |
| Skawinska, Eulalia | Poznan University of Technology | Poland |

| | | |
|---|---|---|
| Spearot, James A. | General Motors Research and Development Center | USA |
| Staretu, Ionel | Transilvania University of Brasov | Romania |
| Sulema, Yevgeniya | National Technical University of Ukraine | Ukraine |
| Suomi, Reima | Turku School of Economics and Business Administration | Finland |
| Suzuki, Junichi | University of Massachusetts | USA |
| Szotek, Sylwia | Wroclaw University of Technology | Poland |
| Tchier, Fairouz | King Saud University | USA |
| Thundat, Thomas | Oak Ridge National Laboratory | USA |
| Tousignant, Michel | Univerity of Sherbrooke | Canada |
| Toussaint, Luis | Universite de Technologie de Belfort Montbeliard | France |
| Trimble, Robert | Indiana University of Pennsylvania | USA |
| Troudt, Edgar E. | CUNY Kingsborough Community College | USA |
| Tsoi, Mai Yin | GGC | USA |
| Valakevicius, Eimutis | Kaunas University of Technology | Lithuania |
| Valdez, Pierre | Georgia Institute of Technology | USA |
| Vasinek, Vladimir | Technical Univerzity of Ostrava | Czech Republic |
| Venkataraman, Satyamurti | All India Association for Micro Enterprise Development | India |
| Vinod, D. S. | Sri Jayachamarajendra College of Engineering | India |
| Wallner, Daniel | Graz University of Technology | Austria |
| Wang, Lei | University of Houston | USA |
| Warwick, Jon | London South Bank University | UK |
| Wellons, Jonathan | Vanderbilt University | USA |
| Whiteley, Rick | Calabash Educational Software | Canada |
| Yadav, S. M. | S.V.National Institute of Technology-Surat | India |
| Yaghmaee, Mohammad Hossien | Ferdowsi University of Mashhad | Iran |
| Yanagisawa, Hideaki | Tokuyama College of Technology | Japan |
| Yingling, Yaroslava | North Carolina State University | USA |
| Yoon, Changwoo | ETRI | South Korea |
| Zalewski, Romuald I. | Poznañ University of Economics | Poland |
| Zaretsky, Esther | Hebrew University | Israel |
| Zelinka, Tomas | Czech Technical University in Prague | Czech Republic |
| Zhu, Hui | Soochow University | China |
| Zobaa, Ahmed | University of Exeter | UK |


# ADDITIONAL REVIEWERS

| | | |
|---|---|---|
| A. Serodio, Carlos M. J. | UTAD University | Portugal |
| Aigner, Werner | University of Linz | Austria |
| Albayrak, Y. Esra | Galatasaray University | Turkey |
| Alhassan, Mohammad | Purdue University Fort Wayne | USA |
| Altan, Metin | Anadolu University | Turkey |
| Andreev, Rumen | Bulgarian Academy of Sciences | Bulgaria |
| Arrabales Moreno, Raúl | Universidad Carlos III de Madrid | Spain |
| Aruga, Masahiro | Teikyo Heisei University | Japan |
| Ashur, Suleiman | Purdue University Fort Wayne | USA |
| Auer, Michael | Carinthia University of Applied Sciences | Austria |
| Badescu, V. | Polytechnic University of Bucharest | Romania |
| Badran, Omar | Al-Balqa` Applied University, Amman | Jordan |
| Balas, Valentina | Aurel Vlaicu University of Arad | Romania |
| Barreiro, J. | Universidad de León | Spain |
| Batovski, Dobri Atanassov | Assumption University of Thailand | Thailand |
| Bayraktar, Seyfettin | Yildiz Technical University | Turkey |
| Bilbao, Josu | IKERLAN-IK4 | Spain |
| Bilich, Feruccio | University of Brasilia | Brazil |
| Botella, Federico | Miguel Hernandez University | Spain |
| Boutejdar, Ahmed | University og Magdeburg Germany | Germany |
| Branzila, Marius | Technical University Gh. Asachi | Romania |
| Burke, Jeffrey | National Pollution Prevention Roundtable | USA |
| Buzzi, Maria Claudia | CNR | Italy |
| Camins, Angel Sevilla | Inbionova Biotech S.L. | Spain |
| Cannavò, Flavio | University of Catania | Italy |
| Cardoso, Pedro | Universidade do Algarve | Portugal |
| Carrió Pastor, M Luisa | Universidad Politécnica de Valencia | Spain |
| Challoo, Linda | Texas A & M University | USA |
| Cheggour, Mohamed | Ecole Normale Superieure (E.N.S.), Marrakech | Morocco |
| Chen, Jing-Heng | Feng Chia University | Taiwan |
| Chen, Ping-Hei | National Taiwan University | Taiwan |
| Chiou , Chuang Chun | Dayeh University | Taiwan |
| Chow, James | University of Toronto | Canada |
| Chu, Hsing-Wei | Lamar University | USA |
| Cotet, Costel Emil | University Politehnica of Bucharest | Romania |
| Csáki, Tibor | University of Miskolc | Hungary |
| D`auria, Francesco | Università Di Pisa | Italy |

| | | |
|---|---|---|
| Dijkstra, Jan | Eindhoven University of Technology | Netherlands |
| Djeffal, Lakhdar | University of Hadj Lakhdar Batna Algeria | Algeria |
| Doughan, Mahmoud | Lebanese University | Lebanon |
| Dougherty, Edmond | Ablaze Development Corp | USA |
| Elwany, Hamdy | Alexandria University | Egypt |
| Fathy, Sherif Kassem | King Fasial University | Saudi Arabia |
| Feldmann, Birgit | University of Hagen | Germany |
| Fernandes de Lima, Milton S. | Institute for Advanced Research (IEAv/CTA) - Brazilian Air Force | Brazil |
| Fidan, Ismail | Tennessee Tech University | USA |
| Fiorentino, Michele | Politecnico di Bari | Italy |
| Fiorini, Rodolfo A. | Politecnico di Milano | Italy |
| Florea, Adrian | Lucian Blaga University | Romania |
| Furtaw, Michael | LI-COR Biosciences | USA |
| Gan, Tat-Hean | TWI Ltd | UK |
| Gandhi, Meenakshi | Guru Gobind Singh Indraprastha University | India |
| Garro, Alfredo | University of Calabria | Italy |
| Ge, Mouzhi | Dublin City University | Ireland |
| Gheno, Simoni | Federal University of Sao Carlos | Brazil |
| Glossman-Mitnik, Daniel | Grupo Nanocosmos | Mexico |
| Glowacki, Miroslaw | University of Science and Technology | Poland |
| Goel, Arun | NIT Kurukshetra | India |
| Goi, Chai Lee | Curtin University of Technology | Malaysia |
| Gonzalez, Emmanuel A. | Jardine Schindler Elevator Corporation | Philippines |
| Gradvohl, Andre L. S. | Universidade Sao Francisco | Brazil |
| Granovsky, Alex | Moscow State University | Russian Federation |
| Grzywaczewski , Marek | Radom Technical University | Poland |
| Gujarathi, Ashish M. | Birla Institute of Technology and Science | India |
| Hajdinjak, Melita | University of Ljubljana | Slovenia |
| Hardman, John | Florida Atlantic University | USA |
| Harris, Marilyn | Capella University | USA |
| Hassini, A. | Laboratory of Application and Analysis of Radiations | Algeria |
| Hassini, Abdelatif | University of Oran Es-Senia | Algeria |
| He, Hongyu | Louisiana State University | USA |
| Hennequin, Sophie | LGIPM | France |
| Hirz, Mario | Graz University of Technology | Austria |
| Hrebicek, Jiri | Masaryk University | Czech Republic |
| Ibhadode, Akii | University of Benin | Nigeria |
| Ingber, Lester | Lester Ingber Research | USA |
| Jaroslav, Heinrich | HBH Projekt Spol. S R.O. | Czech Republic |
| Jarz, Ewald | University of Applied Sciences Kufstein | Austria |
| Jastroch, Norbert | MET Communications | Germany |
| Jegede, Olawale | LM Ericsson (Nigeria) Ltd | Nigeria |
| Jha, Ranjana | Netaji Subhas Institute of Technology | India |
| Jiang, Jinlei | Tsinghua University | China |
| Kaewarsa, Suriya | Rajamangala University of Technology Isan, Sakon Nakhon Campus | Thailand |
| Kakanakov, Nikolay | Technical University of Sofia | Bulgaria |
| Kamrani, Ehsan | Ecole Polytechnique de Montreal | Canada |
| Kässi , Tuomo | Lappeenranta University of Technology | Finland |
| Kawamura, Hidenori | Hokkaido University | Japan |

| | | |
|---|---|---|
| Ke, Bwo-Ren | National Penghu University | Taiwan |
| Khudayarov, Bakhtiyar | TIIM | Uzbekistan |
| Kim, Hyun-Jun | Samsung Electronics | South Korea |
| Kochikar, Vivekanand | Infosys Technologies | India |
| Ksenofontov, Alexandre | Moscow Engineering Physics Institute | Russian Federation |
| Kunold, Ingo | Dortmund University of Applied Sciences and Arts, Institute of Communications Technology | Germany |
| Kureshi, Nadeem | CASE | Pakistan |
| Lee, Joo Hwan | Seoul National University | South Korea |
| Li, Dayong | Shanghai Jiao Tong University | China |
| Li, Yu-Chiang | Southern Taiwan University | Taiwan |
| Lin, Chih-Ting | National Taiwan University | Taiwan |
| Lin, Chun Yuan | Chang Gung University | Taiwan |
| Liu , Tingyang Lewis | National Kaohsiung Normal University | Taiwan |
| Liu, Yajun | South China University of Technology | China |
| Mahdoum, Ali | Centre de Developpement des Technologies Avancees | Algeria |
| Maldonado, J. L. | Centro de Investigaciones en Óptica A.C. | Mexico |
| Mansano, R. D. | Escola Politécnica - USP - PSI - Laboratório de Sistemas Integráveis | Brazil |
| Margelli, F | CNR-ISAC | Italy |
| Markakis, Euaggelos | Technological Educational Institute of Crete | Greece |
| McCormick, John | Institution of Engineering and Technology | UK |
| Medina, Dulce | Universidad Autónoma Metropolitana | Mexico |
| Mierlus-Mazilu, Ion | Technical University of Civil Engineering | Romania |
| Mokhtari, A. | UAV | France |
| Moschim, Edson | State University of Campinas | Brazil |
| Neaga, Elena Iirna | Loughborough University | UK |
| Neves, Filipe | CENIMAT | Portugal |
| Nisar, Humaira | Gwangju Institute of Science and Technology | South Korea |
| Omidvar, Hedayat | Nacional Iranian Gas Company | Iran |
| Onur Hocaoglu, Fatih | Anadolu University | Turkey |
| Pavlidis, George | Research Centre Athena - Cultural and Educational Technology Institute | Greece |
| Pedamallu, Chandra Sekhar | Dept. of Medical Oncology, Dana-Farber Cancer Institute / The Broad Institute | USA |
| Pfliegl, Reinhard | AustriaTech Ltd | Austria |
| Plouffe, B. | Northeastern University | USA |
| Pogarèiæ, Ivan | Polytechnic of Rijeka | Croatia |
| Prakash, S. V. | M.S.Ramaiah Institute of Technology | India |
| Price, Howard | Equipment Design | USA |
| Puslecki, Zdzislaw | Adam Mickiewicz University | Poland |
| Raibulet, Claudia | University of Milano | Italy |
| Raj, Piyush | Goodearth Shipbuilding Pvt Limited / Archean Group | India |
| Rana, Mukhtar Masood | Anglia Ruskin University | USA |
| Ribakov, Y. | Ariel University Center of Samaria | Israel |
| Riesbeck, Christopher | Northwestern University | USA |
| Rodriguez-Florido, M. A. | Canary Islands Institute of Technology | Spain |
| Root, Sara L. | Beckman Coulter | USA |
| Rot, Artur | Wroclaw University of Economics | Poland |
| Rydhagen, Birgitta | Blekinge Institute of Technology | Sweden |
| Sahin, Omer Sinan | Selcuk University | Turkey |

| | | |
|---|---|---|
| Salay Naderi, Mohammad | University of New South Wales | Australia |
| Sánchez , Caio | University of Campinas | Brazil |
| Sane, Vijay | American Chemmical Society | India |
| Sanin, Cesar | The University of Newcastle | Australia |
| Santulli, C. | Universita di Roma Dept. of Chemical Engineering Materials and Environment | Italy |
| Schlue, Benjamin | ARCADIS Consult GmbH | Germany |
| Shannaq, Boumedyen | Nizwa University | Oman |
| Silva, Daniel | Universidade Tiradentes | Brazil |
| Simon, Thierry | LRPmip | France |
| Siricharoen, Waralak V. | University of the Thai Chamber of Commerce | Thailand |
| Sirjani, Mojtaba B. | Norfolk State University | USA |
| Sittidech Siripitayananon, Punnee | Naresuan University | Thailand |
| Skoko, Hazbo | Charles Stuart University | Australia |
| Spina, Edison | Escola Politecnica da Universidade de Sao Paulo | Brazil |
| Šraml, Matjaž | University of Maribor | Slovenia |
| Süral, Haldun | METU | Turkey |
| Sutherland, Trudy | Vaal University of Technology | South Africa |
| Tarng, J. H. | National Chiao Tung University | Taiwan |
| Taylor, Craig | University of Southern California | USA |
| Teodoro, Ana Cláudia | University of Porto | Portugal |
| Tobar, Carlos Miguel | Pontifical Catholic University of Campinas | Brazil |
| Travieso-González, Carlos M. | University of Las Palmas | Spain |
| Trifonova, Silvia | University of National and World Economy | Bulgaria |
| Tseng, Juin-Ling | Minghsin University of Science and Technology | Taiwan |
| Turcu, Cristina | Romanian Society for Automatics and Technical Informatics | Romania |
| Vaganova, Natalia | Institute of Computational Mathematics and Mathematical Geophysics | USA |
| Veeraklaew, Tawiwat | Chulachomklao Royal Military Academy | Thailand |
| Venturini, Mauro | University of Ferrara | Italy |
| Vidal-Naquet, Guy | SUPELEC | France |
| Vimarlund, Vivian | Linköpings Universitet | Sweden |
| Virvilaite, Regina | Kaunas University of Technology | Lithuania |
| Vizureanu, Petrica | Technical University Iasi | Romania |
| Wen, Fuh-Liang | St. John`s University / Taipei Campus | Taiwan |
| William, Gergis | West Virginia University | USA |
| Yazici, Ali | Atilim University | Turkey |
| Yindi, Zhao | China University of Mining and Technology | China |
| Zainal Abidin, Azizan | Universiti Teknologi Petronas | Malaysia |
| Zakrzewska, Danuta | Technical University of Lodz | Poland |
| Zampieri, Douglas E. | UNICAMP | Brazil |
| Zeman, Klaus | Johannes Kepler University Linz | Austria |


# ADDITIONAL REVIEWERS FOR THE NON-BLIND REVIEWING

| | | |
|---|---|---|
| Abdul Raheem, Khalid Fathi | Caledonian College of Engineering- Muscat | Oman |
| Ahuja, Ips | Punjabi University- Patiala | India |
| Albayrak, Y. Esra | Galatasaray University | Turkey |
| Altan, Metin | Anadolu University | Turkey |
| Anchev, Victor | Techical University of Sofia | Bulgaria |
| Arvindkumar, Patel Dilip Kumar | SV National Institute of Technology | India |
| Ataoglu, Senol | University of Texas at Dallas | Turkey |
| Awasthi, Lalit | NIT Hamirpur | India |
| Azambuja, Marcelo | Southern Illinois University | USA |
| B.K.N, Rao | COMADEM | UK |
| Badran, Omar | Al Balqa Applied Science University | Jordan |
| Balagopalan, Sudha | Vidya Academy of Science and Technology | India |
| Barreiro, J. | Universidad de León | Spain |
| Borja, Julio | ITESM Campus Toluca | Mexico |
| Bouslimani, Yassine | University of Moncton | Canada |
| Bubnov, Alexey | Institute of Physics | Czech Republic |
| Byrne, Edmond | University College Cork | Ireland |
| Casolo, Federico | Politecnico di Milano | Italy |
| Challoo, Linda | Texas A & M University | USA |
| Chauhan, Krupesh | S. V. National Institute of Technology | India |
| Chowdhary, Girish | Georgia Institute of Technology | USA |
| Chu, Hsing-Wei | Lamar University | USA |
| Cook, Kristina | Concordia University | Canada |
| Corona Chavez, Alonso | INAOE | Mexico |
| Curreri, Peter A. | NASA | USA |
| Dalto, Edson | IBMEC | Brazil |
| Dean, Vucinic | Vrije Universiteit Brussel (VUB) | Belgium |
| Dijkstra, Jan | Eindhoven University of Technology | Netherlands |
| Dimancescu, Dan | Bastea Enterprises and Holdings LLC | USA |
| Djeffal, Lakhdar | University of Hadj Lakhdar Batna | Algeria |
| Donzelli, Mauro | C.O.N.I. Ferrari | Italy |
| Dougherty, Edward | Ablaze Development Corporation | USA |

| | | |
|---|---|---|
| Dreschhoff, Gisela | University of Kansas | USA |
| Eriksson, Arne | European Commission's Joint Research Centre | Netherlands |
| Ersoy, Sezgin | Marmara Universitesi Teknik Egitim Fakultesi | Turkey |
| Fronte, Daniele | ATMEL | France |
| Gandhi, Meenakshi | Guru Gobind Singh Indraprastha University | India |
| Ghosh, Sadhana | National Institute of Industrial Engineering | India |
| Giberti, Hermes | Politecnico di Milano | Italy |
| Golshan, Mohammad Mehdi | Shiraz university | Iran |
| Gong, Jie | University of Southern Illinois Edwardsville | USA |
| Graff, Mario | Universidad Michoacana de San Nicolás de Hidalgo | Mexico |
| Gutierrez Tornes, Agustin | Citigroup- Banamex and Instituto Politecnico Nacional (IPN) | Mexico |
| Haring, Roland | Ars Electronica | Austria |
| Harris, Marilyn | Capella University | USA |
| Hierlemann, Andreas | Bio Engineering Laboratory (BEL) | Switzerland |
| Hong, Jinglan | Shandong University | China |
| Igarashi, Hajime | Hokkaido University | Japan |
| Ikawa, Yasuo | JAIST | Japan |
| Jaafar, Jafreezal | Universiti Teknologi PETRONAS | Malaysia |
| Jegede, Olawale | LM Ericsson (Nigeria) Ltd | Nigeria |
| JeliÄ‡ Mrcelic, Gorana | Faculty of Maritime Studies | Croatia |
| Jiang, Hui | Stanford University | USA |
| Jilek, Miroslav | Czech Technical University in Prague | Czech Republic |
| Karim, M.A. | Virginia Department of Environmental Quality (DEQ) | USA |
| Khamsi, Amine | The University of Texas at El Paso | USA |
| Khanfar, Khalid | Arabic American University | Israel |
| Knaack, Ulrich | University of Technology Delft | Netherlands |
| Kopustinskas, Vytis | European Commission's Joint Research Centre | Italy |
| Kunold, Ingo | Dortmund University of Applied Sciences and Arts, Institute of Communications Technology | Germany |
| Lang, Werner | Technische Universität München TUM | Germany |
| Laroche, Florent | Ecole Centrale Nantes | France |
| Lee, Tzai-li | National Taiwan Sport University | Taiwan |
| Legendre, Jean-françois | AFNOR | France |
| Lermontov, Mihail | Lermontov | Brazil |
| Liu, Bo | Oak Ridge National Laboratory | USA |
| Liu, Jing Jiang | Zhejiang University | China |
| Liu, Yanfei | Indiana University-Purdue University Fort Wayne | USA |
| Low, Tan Jung | Universiti Teknologi PETRONAS | Malaysia |
| Macwan, J. E. M. | S.V.National Institute of Technology- Surat | India |
| Marpu, Ritu | Georgia Institute of Technology | USA |
| Mayo, Shaker | University of Engineering and Technology- Lahore | Pakistan |
| Morgan, Michael | Liverpool John Moores University | UK |
| Moschim, Edson | State University of Campinas | Brazil |
| Motieifar, Alireza | University of British Columbia | Canada |
| Olubambi, Peter | Tshwane University of Technology | South Africa |
| Omar, Iqbal | Texas A&M University-Kingsville | USA |

| | | |
|---|---|---|
| Onosato, Masahiko | Hokkaido University | Japan |
| Ortega, Felipe | Georgia Institute of Technology | USA |
| Otto, Tauno | Tallinn University of Technology | Estonia |
| Pang, Jian | Changan Automobile Co. Ltd | China |
| Pazienza, Maria Teresa | University Roma 2 | Italy |
| Pfeiffer, Wulf | Fraunhofer-Institut für Werkstoffmechanik (IWM) | Germany |
| Pokorski, Dale | Virginia Tech | USA |
| Prylutskiy, Juriy | Kiyv National University- Kiev- Ukraine | Ukraine |
| Przyby³owski, Piotr | Gdynia Maritime University | Poland |
| Puslecki, Zdzislaw | Adam Mickiewicz University | Poland |
| Raghav, Vrishank | Georgia Institute of Technology | USA |
| Ramos, Antonio | Universidad Michoacana de San Nicolás de Hidalgo | Mexico |
| Richy, Paul | Orange | France |
| Ristov, Panco | Faculty of Maritime Studies | Croatia |
| Robertson, Glen A. | Institute for Advanced Studies in the Space | USA |
| Rodrigues, Jose Alberto | Instituto Superior de Engenharia de Lisboa | Portugal |
| S M, Rizwan | Caledonian College of engineering | Oman |
| Safarik, Pavel | Czech Technical University in Prague | Czech Republic |
| Sanches Silva, Ana Teresa | National Institute of Health Dr. Ricardo Jorge | Portugal |
| Sanin, Cesar | The University of Newcastle | Australia |
| Segalas, Jordi | UNESCO Chair of Sustainability- EPSEVG - UPC | Spain |
| Serhan, Derar | Arizona State University | USA |
| Simsek, Mesut | Yýldýz Teknik Üniversitesi | Turkey |
| Singh, Rupinder | GNDEC | India |
| Singh, Shri Niwas | IIT Kanpur | India |
| Su, J. L. | Tongji University | China |
| Sugasawa, Yoshio | Nihon University of Economy | Japan |
| Tanaka, Kazuyuki | Tohoku University | Japan |
| Tavares, Teresa | University of Minho | Portugal |
| Tomczyk, Andrzej | Rzeszow University of Technology | Poland |
| Variar, Leela | Vidya Academy of Science and Technology | India |
| Vazquez, Manuel | University of Santiago de Compostela | Spain |
| Vesali, Alireza | Furtwangen University | Germany |
| Victor Gopa, Victor | Zaragoza University | Spain |
| Virvilaite, Regina | Kaunas University of Technology | Lithuania |
| Vlachogiannis, Michail | Technological Institute of Larissa | Greece |
| Vollen, Jason | Rensselaer Polytechnic Institute | USA Minor Outlying Islands |
| Wahab, Mohammad | Universiti Kebangsaan | Malaysia |
| Wang, Weiquan | College of Business | China |
| Wen, Zhong | Tsinghua University | China |
| Wu, Yeheng | Case Western Reserve University | USA |
| Yadav, S.M.Yadav | S.V.National Institute of Technology Surat | India |
| Yadav, Kunwar Durg Vijay | S. V. National Institute of Technology | India |
| Yadav, Sanjay | S. V. National Institute of Technology | India |
| Zaneldin, Essam | The United Arab Emirates University | USA |
| Zeman, Klaus | Johannes Kepler University Linz | Austria |

| | | |
|---|---|---|
| Zeng, Qing-An | North Carolina A&T State University | USA |
| Zhao, Xiaoqing | Zhejiang University | China |
| Zhenhua, Zhang | Three Gorges of University | China |

**The 4th International Multi-Conference on Engineering and Technological Innovation: IMETI 2011**


**HONORARY CHAIR**
William Lesso

**PROGRAM COMMITTEE CHAIRS**
Hsing-Wei Chu
C. Dale Zinn

**GENERAL CHAIR**
Nagib Callaos

**ORGANIZING COMMITTEE CHAIRS**
Belkis Sánchez
Andrés Tremante

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


# PROGRAM COMMITTEE

**Chair:** Hsing-Wei Chu (USA)

| | | |
|---|---|---|
| Agrawalla, Raman | Tata Consutlancy Services | India |
| Aguilar, José | Universidad de los Andes | Venezuela |
| Alameh, Kamal E. | Edith Cowan University | Australia |
| Andina, Diego | Technical University of Madrid | Spain |
| Azoulay, J. | Holon Institute of Technology | Israel |
| Barceló, Jaime | Technical University of Cataloni | Spain |
| Bégoin-Augereau, Sandra | University of Poitiers | France |
| Bürkle, Axel | Fraunhofer IOSB | Germany |
| Campbell, D. | Dublin Institute of Technology | Ireland |
| Caron-Pargue, Josiane | University of Poitiers | France |
| Chandersekaran, Coimbatore | Institute for Defense Analyses | USA |
| Chang, Tsun-Wei | Tatung University | Taiwan |
| Colak, Tufan | University of Bradford | UK |
| Cullen, C. | Dublin Institute of Technology | Ireland |
| Delvai, Martin | Technische Universität Wien | Austria |
| Diepold, Klaus | Technische Universität München | Germany |
| Dobosiewicz, W. | University of Guelph | Canada |
| Eckl, Carolin | Technische Universität München | Germany |
| Ellenberg, Robert | Drexel University | USA |
| Fukusako, Takeshi | Kumamoto University | Japan |
| Gréwal, G. | University of Guelph | Canada |
| Halverson, Ranette | Midwestern State University | USA |
| Hamdy, Abeer | Electronics Research Institute | Egypt |
| Hase, William L. | Texas Tech University | USA |
| Horne, Jeremy | RhinoCorps- Ltd. | USA |
| Huang, Yo-Ping | Tatung University | Taiwan |
| Hübner, Dirk | University of Applied Sciences Trier | Germany |
| Jang, Hung-Chin | National Chengchi University | Taiwan |
| Jnifene, Amor | Royal Military College of Canada | Canada |
| Jun, Youngbum | Drexel University | USA |
| Kao, Li-Jen | Tatung University | Taiwan |
| Kim, Moo Wan | Tokyo University of Information Sciences | Japan |
| Kobayashi, Hideo | Mie University | Japan |
| Kousidis, S. | Dublin Institute of Technology | Ireland |
| Kuntz, Noah | Drexel University | USA |
| Labrador, Miguel A. | University of South Florida | USA |
| Leu, Fang-Yie | TungHai University | Taiwan |
| Leung, Lin | BMCC | USA |
| Li, Xiangdong | CUNY | USA |
| Li, Zhi-Yang | TungHai University | Taiwan |
| Liu, Chuan-Ming | National Taipei University of Technology | Taiwan |
| Loh, Robert N. K. | Oakland University | USA |
| López-Nicolás, Carolina | University of Murcia | Spain |

| | | |
|---|---|---|
| Lourderaj, Upakarasamy | Texas Tech University | USA |
| Marchetti, Emanuela | Syddansk University | Denmark |
| Martinez, Christopher | University of New Haven | USA |
| Mbobi, Aime Mokhoo | RedKnee Inc. | Canada |
| McClanahan, Timothy P. | NASA Goddard Space Flight Center | USA |
| McDonnell, C. | Dublin Institute of Technology | Ireland |
| Molina-Castillo, Francisco J. | Univeristy of Murcia | Spain |
| Montabert, Cyril | Virginia Polytechnic Institute and State University | USA |
| Mori, Kazuo | Mie University | Japan |
| Moss, Gloria | University of Glamorgan | UK |
| Müller, Ludìk | University of West Bohemia | Czech Republic |
| Müller, Wilmuth | Fraunhofer IITB | Germany |
| Mundani, Ralf-Peter | Technische Universität München | Germany |
| Musashi, Yasuo | Kumamoto University | Japan |
| Naito, Katsuhiro | Mie University | Japan |
| Nori, Kesav | Tata Consultancy Services | India |
| Oh, Paul | Drexel University | USA |
| Ortwig, Harald | Trier University of Applied Sciences | Germany |
| Overmyer, Scott P. | South Dakota State University | USA |
| Oya, Hidetoshi | Shonan Institute of Technology | Japan |
| Passos, Nelson Luiz | Midwestern State University | USA |
| Piefel, Michael | Humboldt-Universität zu Berlin | Germany |
| Pornthanomwong, Thanawit | Oakland University | USA |
| Portnoy, Jacov | Sami Shamoon College of Engineering | Israel |
| Prashant, Khadka | Texas Tech University | USA |
| Priesler, Miri | Ruppin Academic Center | Israel |
| Pulimeno, Enrico | Università del Salento | Italy |
| Qahwaji, Rami S. R. | University of Bradford | UK |
| Ramos, Ana Luísa | University of Aveiro | Portugal |
| Rank, Ernst | Technische Universität München | Germany |
| Saddler, Audrey M. | University of Technology | Jamaica |
| Sandnes, Frode-Eika | Oslo University College | Norway |
| Sheliga, Michael | Eastern Kentucky University | USA |
| Siddiqui, Muazzam | University of Central Florida | USA |
| Simpson, William R. | Institute for Defense Analyses | USA |
| Su, Meng | The Pennsylvania State University | USA |
| Suárez, Nicolás | Ingeniería y Servicios para la Defensa y el Transporte | Spain |
| Swaminathan, N. | Tata Consultancy Services | India |
| Thuo, Gikiri | Florida Agricultural and Mechanical University | USA |
| Trombka, Jacob I. | NASA Goddard Space Flight Center | USA |
| Uchida, Shoko | Mie University | Japan |
| Valente, Andrea | Aalborg University Esbjerg | Denmark |
| Valle, Andrea | Ericsson Telecomunicazioni | Italy |
| Vasconcelos Ferreira, José | University of Aveiro | Portugal |
| Vaughan, B. | Dublin Institute of Technology | Ireland |
| Visram, Zabin | The University of Warwick | UK |
| Walton, M. | University of Guelph | Canada |
| Williams, Neil | Mississippi State University | USA |
| Xu, Chengying | Purdue University | USA |
| Yanagisawa, Hideaki | Tokuyama College of Technology | Japan |
| Yi, Wang | Dublin Institute of Technology | Ireland |
| Yu, Wen | Instituto Politécnico Nacional | Mexico |
| Zaretsky, Esther | The Academic College for Education Givat Washington | Israel |
| Zhang, Xiaowen | College of Staten Island / CUNY | USA |
| Zhuang, Yu | Texas Tech University | USA |

**8th International Conference on Cybernetics and Information Technologies, Systems and Applications: CITSA 2011**
**in the context of**
**The 4th International Multi-Conference on Engineering and Technological Innovation: IMETI 2011**


# ADDITIONAL REVIEWERS

| | | |
|---|---|---|
| Abe, Jair Minoro | Paulista University | Brazil |
| Abiona, Olatunde | Indiana University Northwest | USA |
| Al. Obaidy, Mohaned | Gulf College | Oman |
| Al-Khassaweneh, Mahmood | Yarmouk University/ Michigan State University | Jordan |
| Azevedo, Isabel | ISEP-IPP | Portugal |
| bajpai, Gaurav | Kigali institute of science and technology | Rwanda |
| Bigan, Cristin | Ecological University of Bucharest | Romania |
| Boutejdar, Ahmed | Universität of MagdeburgEletrical Engineering | Germany |
| Capella Hernández, Juan Vicente | Polytechnical University of Valencia | Spain |
| Caramihai, Mihai | University Politehnica Bucharest | Romania |
| Cuellar, Luis | Austin Community College | USA |
| De Sitter, Jan | University of Antwerp | Belgium |
| Deogun, Jitender | University of Nebraska | USA |
| Eissa, Mahmoud | King Abdulaziz University, Jeddah | Saudi Arabia |
| Erkollar, Alptekin | University of Applied Sciences | Austria |
| Fu, Yujian | Alabama A&M University | USA |
| Gharsallah, Ali | Laboratoire d'Electronique | Tunisia |
| Giannini, Franco | Università di Roma | Italy |
| Guo, Yuan | University of Florida | USA |
| Habib, Maki | The American University in Cairo | Egypt |
| Hoppe, Ulrich | University of Erlangen-NurembergENT-clinic | Germany |
| Horie, Yoshinori | Nihon University | Japan |
| Hossain, Mahmud | Virginia Tech | USA |
| Husak, Miroslav | Czech Technical University in Prague | Czech Republic |
| Imaña, José Luis | Universidad Complutense | Spain |
| Izworski, Andrzej | AGH University of Science and Technology | Poland |
| Jecheva, Veselina | Burgas Free University | Bulgaria |
| Jun, LIANG | Jiangsu University | China |
| Kagan, Albert | Arizona State University | USA |
| Kaneko, Itaru | Tokyo Polytechnic University | Japan |
| Karimi, Bijan | University of New Haven | USA |
| Kinsman, Thomas | Rochester Institute of Technology | USA |
| Koster, Kai | KOSMICON - Koster Media Integration Consulting | Germany |
| Krichene, Assaad | Impact Technologies, LLC | USA |
| La Rue, E | University of Pittsburgh | USA |
| Lam, Vitus | The University of Hong Kong | Hong Kong |
| Lee, In Soo | Kyungpook National University | South Korea |
| Lee, Ki Young | University of Incheon | South Korea |
| Lenuta, Alboaie | S. C. Axiologic Research S. R. L - IT Co. | Romania |

| | | |
|---|---|---|
| Liu, Xiaobo | Edwards and Kelcey | USA |
| Liu, Ziping | Southeast Missouri State University | USA |
| Lo, Wei-Shuo | Meiho Institute of Technology | Taiwan |
| Lu, Ta-Te | Ching Yun University | Taiwan |
| Marinov, Milko | University of Rousse | Bulgaria |
| Mekkakia Maaza, Zoulikha | Université des Sciences et de la Technologie d`Oran | Algeria |
| Miranda, Anibal D. A. | Instituto Tecnológico de Aeronáutica (ITA) | Brazil |
| Misra, Harekrishna | Institute of Rural Management Anand | India |
| Mitloehner, Johann | Vienna University of Economics and Business Administration | Austria |
| Musil, Vladislav | Brno University of Technology | Czech Republic |
| Musirin, Ismail | Universiti Teknologi Mara | Malaysia |
| Naddeo, Alessandro | Università di Salerno | Italy |
| Nickolov, Eugene | National Laboratory of Computer Virology | Bulgaria |
| Nutter, Brian | Texas Tech | USA |
| Oberer, Birgit | University of Applied Sciences Austria | Austria |
| Odetayo, Michael | Coventry University | UK |
| Ong, Soh Khim | National University of Singapore | Singapore |
| Palza, Edgardo | École de Technologie Supérieure | Canada |
| Phokharatkul, Pisit | Mahidol University Salaya | Thailand |
| Rachev, Boris | Technical University of Varna | Bulgaria |
| Raffai, Maria | Szechenyi Istvan University | Hungary |
| Rahimi, Shahram | Southern Illinois University | USA |
| Roehrig, Christof | Univerity of Applied Sciences Dortmund | Germany |
| Romagni, Susana | Universidad Metropolitana | Venezuela |
| Rot, Artur | Wroclaw University of Economics | Poland |
| Schuller, Bjoern | Technische Universitaet Muenchen | Germany |
| Shyr, Wen-Jye | National Changhua University of Education | Taiwan |
| Stapleton, Larry | ISOL Research Centre | Ireland |
| Suh, Koo-Won | Hanyang Cyber University | South Korea |
| Sumari, Putra | Universiti Sains Malaysia | Malaysia |
| Teixeira, Leonor | University of Aveiro | Portugal |
| Tiako, Pierre | Langston University | USA |
| Tsaur, Lih-Feng | Broadcom Inc. | USA |
| Tsaur, Woei-Jiunn | Inderscience Publishers | Taiwan |
| Valova, Irena | University of Rousse | Bulgaria |
| Wang, Wendy | San Jose State University | USA |
| Watanabe, Shigeyoshi | The University of Electro-Communications | Japan |
| Wilkinson, Tim | Cambridge University | UK |
| Wilson, Hentie | University of South Africa UNISA | South Africa |
| Yuan, Shengli | University of Houston - Downtown | USA |
| Zhou, Xiaocong | Sun Yat-Sen University | China |

**8th International Conference on Cybernetics and Information Technologies, Systems and Applications: CITSA 2011**
**in the context of**
**The 4th International Multi-Conference on Engineering and Technological Innovation: IMETI 2011**


# ADDITIONAL REVIEWERS FOR THE NON-BLIND REVIEWING

| | | |
|---|---|---|
| Akorli, Felix | National University of Rwanda | Rwanda |
| Alameh, Kamal E. | Edith Cowan University | Australia |
| Armada, Ana García | Universidad Carlos III de Madrid | Spain |
| Arsov, Silyan | Rousse University | Bulgaria |
| Arsova, Elitsa | University of Rousse "Angel Kanchev" | Bulgaria |
| Bürkle, Axel | Fraunhofer IOSB | Germany |
| Cáceres Reche, Mª Pilar | Researcher of Granada University | Spain |
| Chen, Tai-fang | Providence University | Taiwan |
| Cingula, Marijan | Faculty of Economics and Business Zagreb | Croatia |
| Colak, Tufan | University of Bradford | UK |
| Edoh, Kossi | NC A | USA |
| Garcia, Miguel | University of Ontario Institute of Technology (UOIT) | Canada |
| Georgiev, Tsvetozar | University of Ruse | Bulgaria |
| Gikaru, Wilfred | Egerton University | Kenya |
| Gregory, Mark | Group ESC-Rennes | France |
| Habib, Maki | The American University in Cairo | Egypt |
| Halverson, Ranette | Midwestern State University | USA |
| Hernandez, Orlando | The College of New Jersey | USA |
| Hodjat Kashani, Farokh | IUST | Iran |
| Horne, Jeremy | RhinoCorps- Ltd. | USA |
| Inoue, Yasuhiko | NTT Network Innovation Laboratories | Japan |
| Jamali, Shahram | Moghaddas Ardebili | Iran |
| Jesudoss, Auxeeliya | National University of Rwanda | Rwanda |
| Kabalan, Karim | American University of Beirut | Lebanon |
| Kerbache, Laoucine | HEC Paris | France |
| Kobayashi, Kentaro | Nagoya University | Japan |
| Krichene, Assaad | Impact Technologies, LLC | USA |
| La Rue, E | University of Pittsburgh | USA |
| Liu, Wenyin | University of Hong Kong | Hong Kong |
| Manangi, Suhas | Microsoft India R | India |
| Marinov, Milko | Rousse University | Bulgaria |
| Massiha, Gholam | The University of Louisiana in Lafayette | USA |
| Mizoguchi, Masato | NTT network innovation laboratories | Japan |
| Mohammadkhanli, Leili | Tabriz University | Iran |
| Montabert, Cyril | Virginia Polytechnic Institute and State University | USA |
| Nickolov, Eugene | National Laboratory of Computer Virology | Bulgaria |
| Niyonkuru, Dr. Adronis | National University of Rwanda | Rwanda |
| Ohya, Yuichi | Kansai University | Japan |
| Okada, Hiraku | Saitama University | Japan |
| Pietrobon, Steven Silvio | Small World Communications | Australia |
| Rinaldi, Antonio M. | Università di Napoli Federico II | Italy |
| Su, Jyh-cherng | R.O.C. Military Academy | Taiwan |

| | | |
|---|---|---|
| Syau, Yu-Ru | National Formosa University | Taiwan |
| Twaakyondo, Hashim | University of Daar es slaam | Tanzania |
| Vilardy Perez, Juan M. | Universidad Popular del Cesar | Spain |
| Villaseñor, Luis A. | CICESE Research Centre | Mexico |
| Wang, Walter | Penn State Great Valley | USA |
| Wei, Huaqiang | University of Idaho | USA |
| Yamazato, Takaya | Nagoya University | Japan |
| Yang, King-Jang | Chung Hua University | Taiwan |

**8th International Conference on Cybernetics and Information Technologies, Systems and Applications: CITSA 2011**
**in the context of**
**The 4th International Multi-Conference on Engineering and Technological Innovation: IMETI 2011**


**HONORARY PRESIDENT**
William Lesso

**GENERAL CHAIR**
Michael Savoie

**PROGRAM COMMITTEE CHAIR**
Hsing-Wei Chu

**ORGANIZING COMMITTEE CHAIR**
José Ferrer

**CONFERENCES PROGRAM MANAGER /**
**HARDCOPY PROCEEDINGS PRODUCTION CHAIR**
Maria Sánchez

**TECHNICAL CONSULTANT ON COMPUTING SYSTEMS /**
**CD PROCEEDINGS PRODUCTION CHAIR**
Juan Manuel Pineda

**SYSTEMS DEVELOPMENT, MAINTENANCE AND DEPLOYMENT**
Dalia Sánchez
Keyla Guédez
Nidimar Diaz

**OPERATIONAL ASSISTANT**
Marcela Briceño
Cindi Padilla

**HELP DESK**
Louis Barnes
Sean Barnes
Marisela Jiménez

**META-REVIEWERS SUPPORT**
Maria Sánchez
Dalia Sánchez

**The 2nd International Symposium on Security and Information/Communication Technologies: SICT 2011**
**in the context of**
**The 8th International Conference on Cybernetics and Information Technologies, Systems and Applications: CITSA 2011**
**and**
**The 4th International Multi-Conference on Engineering and Technological Innovation: IMETI 2011**


**GENERAL CHAIR**
Michael Savoie

**ORGANIZING COMMITTEE CHAIR**
José Ferrer

## PROGRAM COMMITTEE

**Chair:** Hsing-Wei Chu (USA)

| | | |
|---|---|---|
| Abusitta, Adel | Ajman University of Science and Technology | United Arab Emirates |
| Alhamouz, Sadeq | Umm Al-Qura University | Saudi Arabia |
| Bica, Ion | Military Technical Academy of Bucharest | Romania |
| Chandramouli, Ramaswamy | National Institute of Standards and Technology | USA |
| Chen, Huan-Ting | National Taiwan University | Taiwan |
| Chen, Pei-Yu | National Taiwan University | Taiwan |
| Coleman, Phillip | Trinity University | USA |
| Czornik, Adam | Silesian University of Technology | Poland |
| Datta, Kamalika | Indian Institute of Tecnology Kharagpur | India |
| Dawoud, Dawoud | University of KwaZulu Natal | South Africa |
| Edwards, Dennis | University of West Florida | USA |
| Fujita, Naoyuki | Japan Aerospace Exploration Agency | Japan |
| Fukase, Masaharu | The University of Tokyo | Japan |
| Furtună, Adrian | Military Technical Academy of Bucharest | Romania |
| Gawliczek, Pawel | Silesian University of Technology | Poland |
| Gordon, Richard | University of KwaZulu Natal | South Africa |
| Karlsson, Kerstin | University of Skövde | Sweden |
| Kowalski, Stewart | Stockholm University | Sweden |
| Kungpisdan, Supakorn | Mahanakorn University of Technology | Thailand |
| Leu, Fang-Yie | TungHai University | Taiwan |
| Li, Zhi-Yang | TungHai University | Taiwan |
| Lin, Frank Yeong-Sung | National Taiwan University | Taiwan |
| Makino, Kohji | SICE | Japan |
| Nawrat, Aleksander | Silesian University of Technology | Poland |
| Nikolova, Evgeniya | Burgas Free University | Bulgaria |
| Nohlberg, Marcus | University of Skövde | Sweden |
| Ramírez Cáceres, Guillermo H. | Soka University | Japan |
| Sato, Tomoaki | Hirosaki University | Japan |
| Sengupta, Indranil | Indian Institute of Technology Kharagpur | India |
| Shuai, Dianxun | East China University of Science and Technology | China |
| Siddiqui, Muazzam | University of Central Florida | USA |
| Simmons, Sharon | University of West Florida | USA |

| | | |
|---|---|---|
| Teshigawara, Yoshimi | Soka University | Japan |
| Thakur, Amrit' Anshu | Mississippi State University | USA |
| Tsang, Po-Hao | National Taiwan University | Taiwan |
| Tsaur, Woei-Jiunn | Da-Yeh University | Taiwan |
| Yamaguchi, Kazunori | The University of Tokyo | Japan |
| Yannakoudakis, E. J. | Athens University of Economics and Business | Greece |
| Zhang, Yu | Trinity University | USA |

## ADDITIONAL REVIEWERS

| | | |
|---|---|---|
| El Manouni, Said | Department of Mathematics, Faculty of Sciences Dhar-Mahraz | Germany |
| Ercan, M. Fikret | Singapore Polytechnic | Singapore |
| Faulin, Javier | Universidad Pública de Navarra | Spain |
| Foina, Aislan | Universidade de SÃ£o Paulo | Brazil |
| Fuster-Sabater, Amparo | Consejo Superior de Investigaciones Científicas | Spain |
| Gao, Meimei | New Jersey Institute of Technology | USA |
| Grouverman, Valentina | RTI International | USA |
| Ho, Liangwei | Collex Communication Corporation | Taiwan |
| Lin, Chun-Wei | National University of Kaohsiung | Taiwan |
| Onwubiko, Cyril | Research Series Limited, London, UK | UK |
| Stoleru, Neculai-Daniel | Commerzbank AG | Germany |
| Thurasamy, Ramayah | Universiti Sains Malaysia | Malaysia |

## ADDITIONAL REVIEWERS FOR THE NON-BLIND REVIEWING

| | | |
|---|---|---|
| Kungpisdan, Supakorn | Mahanakorn University of Technology | Thailand |
| Liu, Chunlei | Valdosta State University | USA |
| Onwubiko, Cyril | Research Series Limited, London, UK | UK |
| Shaffer, Kevin | Fort Hays State University | USA |

**The SUMMER 9th International Conference on Computing, Communications and Control Technologies: CCCT 2011**
**in the context of**
**The 4th International Multi-Conference on Engineering and Technological Innovation: IMETI 2011**


**GENERAL CHAIRS**
Nagib Callaos
Jorge Baralt
Hsing-Wei Chu
Michael J. Savoie

**ORGANIZING COMMITTEE CHAIR**
Belkis Sánchez

# PROGRAM COMMITTEE

**Chairs:** C. Dale Zinn (USA) and José Ferrer (Venezuela)

| | | |
|---|---|---|
| Acharya, Sushil | Robert Morris University | USA |
| Afandizadeh Zargari, Shahriar | Iran University of Science and Technology | Iran |
| Akashi, Takuya | Yamaguchi University | Japan |
| Alam, Hassan | BCL technologies | USA |
| Alameh, Kamal | Edith Cowan University | Australia |
| Anwar, Faisal | Teachers College Columbia University | USA |
| Arikan, Emin | Montclair State University | USA |
| Balasubramanian, Karuppanan | Lefke Avrupa Üniversitesi | Turkey |
| Barkaoui, Kamel | Conservatoire National des Arts et Métiers | France |
| Bradbury, Kyle J. | Tufts University | USA |
| Castaño, Bonifacio | Universidad de Alcalá | Spain |
| Chae, Hui Soo | Teachers College Columbia University | USA |
| Chang, Wen-Kui | Tunghai University | Taiwan |
| Chopra, Deepak | Netaji Subhas Institute of Technology | India |
| Chung, Jen-Yao | IBM | USA |
| Dadji, Yannick | Institute of Computer and Communication Network Engineering | Germany |
| Das, Asesh | Pennsylvania College of Technology | USA |
| Doran, Michael V. | University of South Alabama | USA |
| Drumright-Clarke, Mary Ann | West Virginia University | USA |
| Eshragh, Sepideh | University of Delaware | USA |
| Espinosa Reza, Alfredo | Electrical Research Institute | Mexico |
| Eunni, Madhuri | University of Kansas | USA |
| Faghri, Ardeshir | University of Delaware | USA |
| Fan, Joshua Poh-Onn | University of Wollongong | Australia |
| Fang, Chun-Hsiung | National Kaohsiung University of Applied Sciences | Taiwan |
| Flores, Juan | Universidad Michoacana | Mexico |
| Fukumi, Minoru | University of Tokushima | Japan |
| Galvin, Janine | National Energy Technology Laboratory | USA |
| Gebre, Biruk A. | Stevens Institute of Technology | USA |
| Geyer-Schulz, Andreas | Universität Karlsruhe | Germany |
| Goel, Gaurav | Netaji Subhas Institute of Technology | India |
| Guo, Liwen | Stevens Institute of Technology | USA |
| Gupta, Neha | Maharaja Agrasen Institute of Technology | India |
| Hartono, Rachmat | BCL Technologies | USA |
| Hasnaoui, Salem | École Nationale d'Ingénieurs | Tunisia |
| Hu, Kai-Wei | Tunghai University | Taiwan |
| Iyyunni, Chakradhar | University of Houston | USA |

| | | |
|---|---|---|
| Jenq, John | Monclair State University | USA |
| Katsuyama, Tomoo | Numazu National College of Technology | Japan |
| Kulkarni, Arun | University of Texas | USA |
| Kumar, Aman | BCL Technologies | USA |
| Lalooses, Francine | The MITRE Corporation | USA |
| Leu, Fang-Yie | Tunghai University | Taiwan |
| Lin, Hong | University of Houston | USA |
| Liu, Jun | University of Ulster | UK |
| López-Zafra, Juan-Manuel | Universidad Complutense de Madrid | Spain |
| Low, Kin Huat | Nanyang Technological University | Singapore |
| Maggiore, Giulio | Telecom Itala Mobile | Italy |
| Makaruk, Aleksander | Vienna University of Technology | Austria |
| Mitrovic, Frane | University of Split | Croatia |
| Moosavian, Ali A. | K. N. Toosi University of Technology | Iran |
| Mori, Kinji | Tokyo Institute of Technology | Japan |
| Munitic, Ante | University of Split | Croatia |
| Muñoz, Humberto | Southern University | USA |
| Mura, Gianluca | Politecnico di Milano University | Italy |
| Musser, Jordan | West Virginia University | USA |
| Nagaoka, Tomoyuki | Hoshino Resort Ltd. | Japan |
| Natriello, Gary | Teachers College Columbia University | USA |
| Oya, Hidetoshi | Shonan Institute of Technology | Japan |
| Passos, Nelson | Midwestern State University | USA |
| Patel, Nishit | Stevens Institute of Technology | USA |
| Pochiraju, Kishore | Stevens Institute of Technology | USA |
| Podaru, Vasile | Military Technical Academy | Romania |
| Potente, Till | RWTH Aachen University | Germany |
| Praus, Petr | Charles University | Czech Republic |
| Rahman, Fuad | BCL Technologies | USA |
| Repa, Václav | University of Economics in Prague | Czech Republic |
| Savva, Andreas | University of Nicosia | Cyprus |
| Schlette, Christian | The Institute of Man-Machine Interaction | Germany |
| Seyed-Abbassi, Behrooz | University of North Florida | USA |
| Silva Flores, Brisa Marisol | Electrical Research Institute | Mexico |
| Singh, Madhusudan | Netaji Subhas Institute of Technology | India |
| Smith III, Charles V. | University of South Alabama | USA |
| Srivastava, Smriti | Netaji Subhas Institute of Technology | India |
| Sulc, Bohumil | Czech Technical University | Czech Republic |
| Suomi, Reima | Turku School of Economics and Business Administration | Finland |
| Tanaka, Kanya | Yamaguchi University | Japan |
| Tarnikova, Yuliya | BCL Technologies | USA |
| Treven, Sonja | University of Maribor | Slovenia |
| Ucan, Osman N. | Istanbul University | Turkey |
| Visram, Zabin | University of Warwick | UK |
| Vrána, Stanislav | Czech Technical University in Prague | Czech Republic |
| Yao, Wenbing | Brunel University | UK |
| Yazawa, Toru | Tokyo Metropolitan University | Japan |
| Zolotová, Iveta | Technical University in Kosice | Slovakia |

## ADDITIONAL REVIEWERS

| | | |
|---|---|---|
| Aldasht, Mohammed | Palestine Polytechnic University | Palestine |
| Almachi, Blanca | | Ecuador |
| Averweg, Udo | Durban Institute of Technology | South Africa |
| Cheng, Chien-Fu | National Chiao Tung University | Taiwan |
| Ferronato, Pierfranco | Soluta.Net | Italy |
| Gini, Giuseppina | Politecnico di Milano | Italy |
| Hanakawa, Noriko | Hannan University | Japan |
| Hlynka, Myron | University of Windsor | Canada |
| Katos, Vasilios | Democritus University of Thrace | Greece |
| Khmelevsky, Youry | Okanagan College | Canada |
| Kim, Dong W. | INHA Technical College | South Korea |
| Machado, José | Institute of Engineering of Porto | Portugal |
| Ng, Kia | University of Leeds | UK |
| Nleya, Bakhe | University of Natal | South Africa |
| Parrilla, Luis | University of Granada | Spain |
| Prykarpatsky, Anatoliy | AMM University of Science and Technology | Poland |
| Röning, Juha | University of Oulu | Finland |
| Saxena, Manoj | University of Delhi, Karampura | India |
| Sembera, Jan | Technical University of Liberec | Czech Republic |
| Sheng, Ye | University of California Berkeley | USA |
| Usrey, Michael | Energy Window, Inc. | USA |
| Zaretsky, Esther | Giv`at Washington Academic College of Education | Israel |
| Zhou, Zehai | University of Houston-Downtown | USA |

## ADDITIONAL REVIEWERS FOR THE NON-BLIND REVIEWING

| | | |
|---|---|---|
| Abd Wahab, Mohd Helmy | Universiti Tun Hussein Onn Malaysia | Malaysia |
| Baek, Jinsuk | Winston-Salem University | USA |
| Edoh, Kossi | North Carolina Agricultural and Technical State University | USA |
| Gutierrez, Angel | Montclair State University | USA |
| Junnapiya, Somyot | KMITL | Thailand |
| Kobe, Sigismund | Technische Universität Dresden | Germany |
| Moungnoul, Phichet | KMITL | Thailand |
| Prato, Luca | Politecnico di Milano | Italy |
| Schürmann, Volker | University of Applied Sciences Bochum | Germany |
| Unruh, William | UBC | Canada |

**The 3rd International Symposium on Optical Engineering and Photonic Technology: OEPT 2011**
**in the context of**
**The 4th International Multi-Conference on Engineering and Technological Innovation: IMETI 2011**


**GENERAL CHAIR**
Nagib Callaos

**ORGANIZING COMMITTEE CHAIRS**
Belkis Sánchez
Andrés Tremante

## PROGRAM COMMITTEE

**Chairs:** Hsing-Wei Chu (USA)
C. Dale Zinn (USA)

| | | |
|---|---|---|
| Abusitta, Adel | University of Jordan | United Arab Emirates |
| Aoki, Toru | Shizuoka University | Japan |
| Carbon, Margarita | The Boeing Co | USA |
| Elmahboub, W. M. | Hampton University | USA |
| Gofron, Kazimierz J. | Argonne National Laboratory | USA |
| Jivkova, Svetla | Pennsylvania University State University | USA |
| Lin, Shih-Hung | National Taiwan University | Taiwan |
| Marsalek, Leos | Technical University of Ostrava | Czech Republic |
| Morii, Hisashi | Shizuoka University | Japan |
| Nakashima, Takuya | Shizuoka University | Japan |
| Neo, Yoichiro | Shizuoka University | Japan |
| Schmitt, Robert | Fraunhofer Institute for Production Technology | Germany |
| Stubberud, Stephen | The Boeing Company | USA |
| Stürwald, Stephan | Fraunhofer Institute for Production Technology | Germany |
| Tornow, Michael | University Magdeburg | Germany |
| Vasinek, Vladimir | Technical Univerzity of Ostrava | Czech Republic |
| Zhao, Yequan | Harbin Institute of Technology | China |

## ADDITIONAL REVIEWERS

| | | |
|---|---|---|
| Barry, Bazara | University of Khartoum | Sudan |
| Bitter, Manfred | Princeton University, Princeton Plasma Physics Laboratory | USA |
| Bouslimani, Yassine | University of Moncton | Canada |
| Cao, Hongjun | D-Tech Optoelectronics, Inc. | USA |
| Capraro, Ivan | University of Padua | Italy |
| Carvalho, Joel Pedro | INESC Porto Optoelectronics and Electronic Systems Unit | Portugal |
| Celenk, Mehmet | Ohio University | USA |
| Efremov, Mikhail | Institute of Semiconductor Physics | Russian Federation |
| Fiorini, Rodolfo | Politecnico di Milano | Italy |
| Guo, Huanqing | National University of Ireland, Galway | Ireland |
| Ho, Wen-Hsien | Kaohsiung Medical University | Taiwan |

| | | |
|---|---|---|
| Hussain, Aini | Institute of Electrical and Electronics Engineers | Malaysia |
| Iyengar, Satish | Syracuse University | USA |
| Jacinto, Carlos | Federal University of Alagoas | Brazil |
| Joshi, Ravi | Old Dominion University | USA |
| Konidaris, Nicholas | California Institute of Technology | USA |
| Kuo, Wen-Chuan | National Taiwan Normal University | Taiwan |
| Lebigot, Eric | Laboratoire Kastler Brossel (Ecole Normale Supérieure, CNRS, Université P. | France |
| Lee, Jen-Ai | Taipei Medical University | Taiwan |
| Li, Ming | GE Global Research | USA |
| Liang, Baolai | California Nanosystems Institute (CNSI), University of California, Los Angeles (UCLA) | USA |
| Liu, HsinChih Frank | Shih-Chien University | Taiwan |
| Magne, Sylvain | CEA | France |
| Majji, Manoranjan | Texas A & M University | USA |
| Malacara-Doblado, Daniel | Centro de Investigaciones en Optica, A. C. | Mexico |
| Malomed, Boris | Tel Aviv University | Israel |
| Martiradonna, Luigi | National Nanotechnology Laboratory of CNR/INFM | Italy |
| Mersmann, Christoph | WZL RWTH | Germany |
| Moritz, Tobias | University of California, San Francisco | USA |
| Mosorov, Volodymyr | Technical University of Lodz | Poland |
| Nagy, Norbert | Research Institute for Technical Physics and Materials Science - MFA | Hungary |
| Nourrit, Vincent | The University of Manchester | UK |
| Paz-Viera, Juan | Universidad Central Marta Abreu de las Villas | Cuba |
| Pham, Phuong-T | Penn State University | USA |
| Piatti, Dario | Politecnico di Torino | Italy |
| Piron, Vianney | ENSAM | French Guiana |
| Pujol Baiges, Maria Cinta | Universitat Rovira i Virgili | Spain |
| Richard, Johan | ICC / Durham University | UK |
| Saeki, Soichi | Applied Medical Engineering Science, Graduate School of Medicine, Yamaguchi University | Japan |
| Sandeau, Nicolas | Université Pierre et Marie Curie Paris VI | France |
| Sandoval-Romero, Gabriel E. | Centro de Ciencias Aplicadas y Desarrollo Tecnologico-Universidad Nacional Autonoma de Mexico | Mexico |
| Scalise, Lorenzo | Università Politecnica delle Marche | Italy |
| Selby, Richard | Czech University of Life Sciences | Czech Republic |
| Sirbu, Dan | Princeton University | Canada |
| Small, Alexander | California State Polytechnic University, Pomona | USA |
| Song, Renbo | Lehigh University | USA |
| Sun, Hongyue | University of Aberdeen | UK |
| Suzuki, Kenji | University of Chicago | USA |
| Szabo, Csilla | U. S. Naval Research Laboratory | USA |
| Tian, Fenghua | Bioengineering Department, University of Texas at Arlington | USA |
| Tong, Xiaojun | Harbin institute of technology, Weihai | China |
| Velotta, Raffaele | University of Naples | Italy |
| Vermiglio, Giuseppe | Universitï¿½ degli Studi di Messina | Italy |
| Watekar, Pramod | VIT University, Chennai | India |
| Wortman, Dirk | RWTH Aachen University | Germany |
| Yan, Zhiqing | Eastern Illinois University | USA |
| Yang, Wei | University of Ottawa | Canada |

| | | |
|---|---|---|
| Yao, Xincheng | University of Alabama at Birmingham | USA |
| Yoder, Elizabeth | University of California | USA |
| Yuen, Peter | Cranfield University | UK |
| Yusoff, Zulfadzli | Multimedia University | Malaysia |
| Zhang, Bo | OPNEXT | USA |

## ADDITIONAL REVIEWERS FOR THE NON-BLIND REVIEWING

| | | |
|---|---|---|
| Alameh, Kamal | Electron Science Research Institute- Edith Cowan University | Australia |
| Chow, Yuk-Tak | City University of Hong Kong | Hong Kong |
| Deluca, Jean-Pierre | Alcatel Vacuum Products | USA |
| Kennedy, Brendan | Optical Biomedical Engineering Laboratory- University of Wes | Australia |
| Kothari, Dilip Kumar | Nirma University | India |
| Li, Wen-Tyng | Chung Yuan Christian University | Taiwan |
| Pathak, S. S. | Indian Institute of Technology- Karagpur | India |
| Peng, Gang Ding | University of New South Wales | Australia |
| Rastogi, Vipul | Indian Institute of Technology- Roorkee | India |
| Refahi Oskouei, Amir | Amirkabir University of Technology | Iran |
| Refregier, Philippe | Institut Fresne | France |
| Rudge Barbosa, Felipe | Universidade Estadual de Campinas - Unicamp | Brazil |
| Sánchez-Pérez, Celia | Centro de Ciencias Aplicadas y Desarrollo Tecnológico, Universidad Nacional Autónoma de México, UNAM | Mexico |
| Singh, Surya Pal | Institute of Technology- Banaras Hindu University | India |
| Sun, Jing | Zhongshan (Sun Yat-Sen) University | China |
| Wolff, Markus | HAW Hamburg | Germany |
| Wu, Wan | Science Systems and Applications, Inc | USA |
| Yang, Shiyuan | Kyushu Institute of Technology | Japan |
| Yasumoto, Kiyotoshi | Kyushu University | Japan |
| Zeamer, Michelle | Abbess Instruments | USA |
| Zhan, Qiwen | University of Dayton | American Samoa |
| Zhu, Bin | Duke University | USA |

**The 3rd International Symposium on Energy Engineering, Economics and Policy: EEEP 2011**
**in the context of**
**The 4th International Multi-Conference on Engineering and Technological Innovation: IMETI 2011**

**GENERAL CHAIRS**
Andrés Tremante
Nagib Callaos (IMETI)

**ORGANIZING COMMITTEE CHAIRS**
Jorge Baralt
Friedich Welsch
Belkis Sánchez

## PROGRAM COMMITTEE

**Chairs:** Freddy Malpica (Venezuela)
Thierry Lefevre (Thailand)

| | | |
|---|---|---|
| Akinnikawe, Oyewande | Texas A&M University | USA |
| Anchliya, Abhishek | Texas A&M University | USA |
| Arena, Umberto | Second University of Naples | Italy |
| Baruah, Debendra Chandra | Tezpur University | India |
| Basso, Giuliano | IEEE | Belgium |
| Brezet, J. C. | Delft University of Technology | Netherlands |
| Carrasquero, José Vicente | Universidad Simón Bolívar | Venezuela |
| Cázares-Rangel, Víctor M. | UANL-ITESM-UEZ | Mexico |
| Cerny, Vaclav | University of West Bohemia | Czech Republic |
| Cha, Seung Tae | Korea Electric Power Research Institute | South Korea |
| Ehlig-Economides, Christine | Texas A&M University | USA |
| Elkamel, Ali | University of Waterloo | Canada |
| Fukase, Masa-Aki | Hirosaki University | Japan |
| Gustavsson, Rune | Blekinge Institute of Technology | Sweden |
| Kan, S. Y. | Delft University of Technology | Netherlands |
| Kim, Yong Hak | Korea Electric Power Research Institute | South Korea |
| Klapp, Jaime | Instituto Nacional de Investigaciones Nucleares | Mexico |
| Lefevre, Thierry | Center for Energy Environment Resources Development | Thailand |
| Mastellone, Maria Laura | Second University of Naples | Italy |
| Melioli, Riccardo | Università Degli Studi di Genova | Italy |
| Naddeo, Alessandro | Università di Salermo | Italy |
| Platt, Glenn | CSIRO Energy Technology | Australia |
| Rahman, Anuar Abdul | Pusat Tenaga Malaysia | Malaysia |
| Revetria, Roberto | Università Degli Studi di Genova | Italy |
| Riaz Moghal, Mohammad | University College of Engineering and Technology | Pakistan |
| Shin, Jeong Hoon | Korea Electric Power Research Institute | South Korea |
| Sues, Anna | Eindhoven University of Technology | Netherlands |
| Tam, Wing K. | Swinburne University of Technology | Australia |
| Tremante, Andrés | Universidad Simón Bolívar | Venezuela |

| | | |
|---|---|---|
| Velte, Clara Marika | Technical University of Denmark | Denmark |
| Zaccariello, Lucio | Second University of Naples | Italy |
| Zobaa, Ahmed | Cairo University | UK |

**ADDITIONAL REVIEWERS**

| | | |
|---|---|---|
| Al-Ammar, Essam | King Saud University | Saudi Arabia |
| Arsov, Silyan | University of Ruse | Bulgaria |
| Barzev, Kiril | University of Ruse | Bulgaria |
| Christian, Jeff | Lawson Lundell LLP | Canada |
| Connolly, James | Nanophotonics Centre, Universidad Politecnica de Valencia | Spain |
| Crasto, Giorgio | WindSim | Norway |
| Dassiou, Xeni | City University | UK |
| De Benedictis, Michele | Politecnico di Bari | Italy |
| Ferrando, Emanuele | Selex Galileo | Italy |
| Fiorini, Rodolfo | Politecnico di Milano | Italy |
| Graditi, Giorgio | ENEA, the Italian National Agency for New Technologies, Energy and the Environment | Italy |
| Hammad, Mahmoud | DG. of Global Network of Renewable Energy Approaches for Desert Regions | Jordan |
| Jayaraj, Simon | NIT Calicut | India |
| Jha, Ranjana | Netaji Subhas Institute of Technology | India |
| Khanna, Rita | The University of New South Wales, Sydney | Australia |
| Khatib, Hisham | World Energy Council | Jordan |
| Lakhoua, Mohamed Najeh | University of Carthage | Tunisia |
| Ma, Jian | Pacific Northwest National Laboratory | USA |
| Ma, Wenping | Center for Applied Energy Research, University of Kentucky | USA |
| Mensah-Biney, Robert | NCSU Minerals Research Laboratory | USA |
| Muensit, Nantakan | Prince of Songkla University | Thailand |
| Owen, Alan | Robert Gordon University | UK |
| Pimentel, David | Cornell University | USA |
| Poggi, Philippe | University of Corsica | France |
| Schenk, Peer | The University of Queensland | Germany |
| Singh, Kaushlendra | West Virginia University | USA |
| Vale, Zita | ISEP / GECAD Institute of Engineering - Polytechnic of Porto | Portugal |
| Yeh, Sonia | University of California, Davis | USA |
| Zhao, Jinquan | Hohai University | China |

**ADDITIONAL REVIEWERS FOR THE NON-BLIND REVIEWING**

| | | |
|---|---|---|
| Arnold, Johannes | German Aerospace Center | Germany |
| Bessler, Wolfgang | German Aerospace Center | Germany |
| Clemente, Frank | The Pennsylvania State University | USA |
| Criqui, Patrick | CNRS | France |
| Doujak, Eduard | Vienna University of Technology | Austria |
| Escudero, Ana Cecilia | Universidad Pontificia Bolivariana | Colombia |
| Ferrando, Emanuele | Selex Galileo | Italy |
| Flocard, Hubert | CNRS | France |
| Forsell, Nicklas | Mines Paristech | France |
| Guerassimoff, Gilles | MINES ParisTech | France |
| Haase, Marek | Czech Coal | Czech Republic |
| Hass, Reinhard | Vienna University of Technology | Austria |
| Icil, Huriye | Eastern Mediterranean University | Cyprus |
| Olaya, Yris | Universidad Nacional de Colombia | Colombia |
| Pellegrino, Michele | ENEA, the Italian National Agency for New Technologies, Energy and the Environment | Italy |
| Selosse, Sandrine | Centre for Applied Mathematics- MINES ParisTech | France |
| Tam, Wing K. | Swinburne University of Technology | Australia |
| Wagner, Norbert | German Aerospace Center | Germany |
| Yang, Qiujing | The Pennsylvania State University | USA |

**3rd International Symposium on Engineering Education and Educational Technologies: EEET 2011**
**in the context of**
**The 4th International Multi-Conference on Engineering and Technological Innovation: IMETI 2011**


**GENERAL CHAIR**
Nagib Callaos

**ORGANIZING COMMITTEE CHAIRS**
Belkis Sánchez
Andrés Tremante

## PROGRAM COMMITTEE

**Chairs:** Hsing-Wei Chu (USA)
Jorge Baralt (Venezuela)
José Vicente Carrasquero (Venezuela)

| | | |
|---|---|---|
| Acharya, Sushil | Robert Morris University | USA |
| Adoghe, Loretta | Miami Dade College | USA |
| Airola, Denise | University of Arkansas | USA |
| Alba-Castro, José L. | Universidad de Vigo | Spain |
| Alcina, M. Amparo | Universitat Jaume I | Spain |
| Aldana Segura, Waleska | Universidad de San Carlos de Guatemala | Guatemala |
| Alshara, Osama | Higher Colleges of Technologies | United Arab Emirates |
| Anderson, Travis | University of Arkansas | USA |
| Argones-Rúa, Enrique | Universidad de Vigo | Spain |
| Azaiza, Khitam | Nova Southeastern University | USA |
| Barranco-Mendoza, Alma | Trinity Western University | Canada |
| Belcher, E. Christina | Trinity Western University | Canada |
| Bennett, Leslie | University of Louisville | USA |
| Blatt, Inge | University of Hamburg | Germany |
| Blignaut, Seugnet | North-West University | South Africa |
| Boeller, Nadja | University of Applied Sciences | Switzerland |
| Bos, Wilfried | Dortmund University of Technology | Germany |
| Burke, David | Robert Morris University | USA |
| Burnett, Andrea | University of the West Indies | Barbados |
| Caileanu, Corneliu | Technical University of Iasi | Romania |
| Call, Anson B. | Iowa State University | USA |
| Carswell, Andrew T. | University of Georgia | USA |
| Chen, Kuo Lane | University of Southern Mississippi | USA |
| Chen, Li-Ling | California State University | USA |
| Chia, Rosina | East Carolina University | USA |
| Chiou, Richard | Drexel University | USA |
| Cubukcu, Feryal | Dokuz Eylul University | Turkey |
| Dampier, David A. | Mississippi State University | USA |
| Darabi Golshani, Ali M. | Curtin University of Technology | Australia |
| Dark, Melissa J. | Purdue University | USA |
| Debela , Nega | Marshall University | USA |
| Devlin, Marie | Newcastle University | UK |
| Digangi, Samuel | Arizona State University | USA |
| Dosi, Vasiliki | University of Ioannnina | Greece |
| Drummond, Sarah | Durham University | UK |

| | | |
|---|---|---|
| Duncan, Steve | East Carolina University | USA |
| Dunning, Jeremy | Indiana University | USA |
| Dunning, Katherine | ITT Technical Services | USA |
| Edmondson, James R. | Vanderbilt University | USA |
| Edwards, Stephen H. | Virginia Tech | USA |
| Els, Christo | North-West University | South Africa |
| El-Sheikh, Eman M. | University of West Florida | USA |
| Eye, John | Southern Utah University | USA |
| Fernandes, Márcia | University of Uberlândia | Brazil |
| Fisher, Wendy | The Open University | UK |
| Freeman, Roger | University of Washington | USA |
| Frizell, Sherri S. | Prairie View AM University | USA |
| Galloway, Marvin | Northwest Arkansas Community College | USA |
| Gaspar, Alessio | University of South Florida | USA |
| Gendron, Dennis | Tennessee State University | USA |
| Godwin, Clark | University of South Florida | USA |
| Göl, Özdemir | University of South Australia | Australia |
| González-Agulla, Elisardo | Universidad de Vigo | Spain |
| Goodwin, Dave | National Energy Research Scientific ComputingCenter | USA |
| Goulding, Tom | Daniel Webster College | USA |
| Goy, Martin | Dortmund University of Technology | Germany |
| Gravvanis, George | Democritus University of Thrace | Greece |
| Green, Shelley | Nova Southeastern University | USA |
| Greener, Susan L. | University of Brighton | UK |
| Guijarro-Fuentes, Pedro | University of Plymouth | UK |
| Hall, Cathy | East Carolina University | USA |
| Hall, Tom L. | TCC Publishing, Inc. | USA |
| Hammam, Mahmoud | Helwan University | Egypt |
| Hart, Alexis | Virginia Military Institute | USA |
| Hasnaqui, Salem | École Nationale d'Ingénieurs de Tunis | Tunisia |
| Hatch, Andrew | Durham University | UK |
| Hendel, Russell Jay | Towson University | USA |
| Herget, Josef | University of Applied Sciences | Switzerland |
| Herrnstadt, Steven | Iowa State University | USA |
| Hierl, Sonja | University of Applied Sciences | Switzerland |
| Hodge, Diane M. | Radford University | USA |
| Holifield, David | University of Wales Institute Cardiff | UK |
| Hsieh, Kun-Lin (Johnny) | Nanhua University | Taiwan |
| Hsu, Mei-Hua | Chang Gung Institute of Technology | Taiwan |
| Hudson, Clemente Charles | Valdosta State University | USA |
| Ito, Akinori | Tohoku University | Japan |
| Izydorczyk, Jacek | Silesian University of Technology | Poland |
| Jannasch-Pennell, Angel | Arizona State University | USA |
| Jara Guerrero, Salvador | University of Michoacan | Mexico |
| Jomhari, Nazean | Universiti Malaya | Malaysia |
| Jones, Lovell | University of Texas | USA |
| Ju, Teresa L. | University of Science and Technology | Taiwan |
| Kang, Seok | Arizona State University | USA |
| Karamat, Parwaiz | The Open Polytechnic of New Zealand | New Zealand |
| Katz, Irvin R. | Educational Testing Service | USA |
| Kim, Chang | Arizona State University | USA |
| Kimmins, Dovie L. | Middle Tennessee State University | USA |
| King, Denae | M.D Anderson Cancer Center | USA |
| Kizirian, Robin | Drexel University | USA |

| | | |
|---|---|---|
| Klosowski, Piotr | Silesian University of Technology | Poland |
| Knowles, LaSonya | University of Texas | USA |
| Kolar, Josef | Czech Technical University in Prague | Czech Republic |
| Konno, Tatsuki | Tohoku University | Japan |
| Koshy, Swapna | Univerity of Wollongong in Dubai | United Arab Emirates |
| Koutselini, Mary | University of Cyprus | Cyprus |
| Krakowska, Monika | Jagiellonian University | Poland |
| Kraska, Lena | University of Hamburg | Germany |
| Kurubacak, Gulsun | Anadolu University | Turkey |
| Kwon, Yongjin James | Ajou University | South Korea |
| Lawless, Kimberly A. | University of Illinois | USA |
| Lee, Huei | Eastern Michigan University | USA |
| Letia, Tiberiu | Technical University of Cluj-Napoca | Romania |
| Lewis, Anissa | University of Texas | USA |
| Li, Jingyi | University of Maryland | USA |
| Lingard, Robert W. | California State University- Northridge | USA |
| Liu, Eric Zhi-Feng | National Central University | Taiwan |
| Livne, Nava L. | University of Utah | USA |
| Livne, Oren E. | University of Utah | USA |
| Lizano-DiMare, Maria | Sacred Heart University | USA |
| Lo, WenJuo | Arizona State University | USA |
| Long, Lori | Arizona State University | USA |
| Lowe, John | University of Bath | UK |
| Lowry, Pam | Lawrence Technological University | USA |
| Machotka, Jan | University of South | Australia |
| Macklin, Alexius Smith | Purdue University | USA |
| Manigault, Carol W. | Purdue University | USA |
| Mansour, Samah | Grand Valley State University | USA |
| Marshall, Susanne | Nova Southeastern University | USA |
| Martin, Mary B. | Middle Tennessee State University | USA |
| Martinovic, Dragana | University of Windsor | Canada |
| McCreery, Michael | University of Nevada | USA |
| McKenzie, Sarah | University of Arkansas | USA |
| Mehrabian, Ali | University of Central Florida | USA |
| Milgrom, Tzachi | Hadassah College Jerusalem | Israel |
| Moore, Karla Alvarado | University of Central Florida | USA |
| Mulvenon, Sean W. | University of Arkansas | USA |
| Nahmens, Isabelina | University of Central Florida | USA |
| Nave, Felecia M. | Prairie View A & M University | USA |
| Nedic, Zorica | University of South Australia | Australia |
| Nedunchelian, Ramanujam | Sri Venkateswara College of Engineering | India |
| Nikraz, Hamid | Curtin University of Technology | Australia |
| Olla, Phillip | Madonna University | USA |
| Parker, Brenda C. | Middle Tennessee State University | USA |
| Peerani, Naveed | Nova Southeastern University | USA |
| Pfeifer, Michael | Dortmund University of Technology | Germany |
| Phillips, C. Dianne | Northwest Arkansas Community College | USA |
| Phillips, E. Ray | Middle Tennessee State University | USA |
| Pierce, Tamyra A. | California State University | USA |
| Plakitsi, Katerina | University of Ioannina | Greece |
| Prescott, David | Universiti Brunei Darussalam | Brunei Darussalam |
| Quershi, Elena | Madonna University | USA |
| Rajamony, Bhuvenesh | University Malaysia Perlis | Malaysia |
| Ralevich, Victor | Sheridan College | Canada |
| Richards, Louise | Leeds Met University | UK |

| | | |
|---|---|---|
| Riggs, Ken R. | FAMU | USA |
| Rodi, Anthony | California University of Pennsylvania | USA |
| Rude-Parkins, Carolyn | The University of Louisville | USA |
| Rustad, Julie | The College of St. Scholastica | USA |
| Rutherfoord, Rebecca H. | Southern Polytechnic State University | USA |
| Ryel, Regina | NorthWest Arkansas Community College | USA |
| Sandnes, Frode Eika | Oslo University College | Norway |
| Sanger, Patrick | Alvin Community College | USA |
| Schrader, P. G. | University of Nevada | USA |
| Sharma, Narendra | Naugatuck Valley Community College | USA |
| Sharp, Simon | Leeds Metropolitan University | UK |
| Shing, Chen-Chi | Radford University | USA |
| Shing, Marn-Ling | Taipei Municipal University of Education | Taiwan |
| Sico, John | Nova University | USA |
| Smith, Derek | Victoria University of Wellington | New Zealand |
| Soeiro, Alfredo | University Porto | Portugal |
| Stanaback, Joe | University of South Florida | USA |
| Stay, Victoria | Arizona State University | USA |
| Stegman, Charles | University of Arkansas | USA |
| Stewart, Gay | University of Arkansas | USA |
| Stvan, Laurel | University of Texas Arlington | USA |
| Su, Hui Fang Huang | Nova Southeastern University | USA |
| Suzuki, Motoyuki | Tohoku University | Japan |
| Swart, William | East Carolina University | USA |
| Tan, Roy Patrick | Virginia Tech | USA |
| Taylor, Stephen | Sussex University | UK |
| Thirunarayanan, M. O. | Florida International University | USA |
| Thornton, Matthew | Virginia Tech | USA |
| Torkul, Orhan | Üniversitesi Sakarya | Turkey |
| Traum, Maria | Johannes Kepler University | Austria |
| Tseng, Tzu-Liang Bill | The University of Texas | USA |
| Tsiriga, Victoria | University of Piraeus | Greece |
| Ureel II, Leo C. | Northwestern University | USA |
| Vaca, Roberto | Sanger High School | USA |
| Vaida, Mircea-Florin | Technical University of Cluj-Napoca | Romania |
| Voss, Andreas | Dortmund University of Technology | Germany |
| Walters, Sharon E. | Southern Illinois University | USA |
| Wang, Kening | University of Arkansas | USA |
| Wells, Harvey | Institute of Psychiatry- King`s College London | UK |
| Welsch, Friedrich | Universidad Simón Bolívar | Venezuela |
| Wight, Charles A. | University of Utah | USA |
| Williams, Greg | University of Maryland | USA |
| Williams van Rooij, Shahron | George Mason University | USA |
| Xeni, Elena | University of Cyprus | Cyprus |
| Yang, Yueh-Ting | Drexel University | USA |
| Yu, Xin | University of Bath | UK |
| Zainon Hamzah, Zaitul Azma | Universiti Putra Malaysia | Malaysia |
| Zitman, Taylor | Indiana University | USA |
| Zogla, Irena | University of Latvia | Latvia |
| Zwaneveld, Bert | Open University | Netherlands |

## ADDITIONAL REVIEWERS

| | | |
|---|---|---|
| Abuzir, Yousef | AlQuds Open University | West Bank |
| Dampier, David | Mississippi State University | USA |
| El Kashlan, Ahmed | Academy for Science and Technology | Egypt |
| Florescu, Gabriela | Institute National de C-D în Informaticã | Romania |
| Fougeres, Alain-Jerome | Université de Technologie de Belfort-Montbéliard | France |
| Frosch-Wilke, Dirk | University of Applied Sciences | Germany |
| Guasch, Teresa | Open University of Catalonia (UOC) | Spain |
| Potorac, Alin Dan | University of Suceava | Romania |
| Prodan, Augustin | Iuliu Hatieganu University | Romania |
| Ritter, Arthur | Stevens Institute of Technology | USA |
| Sherman, Barbara Ann | Buffalo State College | USA |
| Uddin, Mueen | Universiti Teknologi Malaysia | Malaysia |
| Wang, Ching-Huang | National Formosa University | Taiwan |
| Xie, Haiyan | University of Arkansas at Little Rock | USA |

## ADDITIONAL REVIEWERS FOR THE NON-BLIND REVIEWING

| | | |
|---|---|---|
| Alberts, Maris | Institute of Mathematics and Computer Science | Latvia |
| Alsop, Graham | Kingston University | UK |
| Burian, Jaroslav | Palacky University in Olomouc | Czech Republic |
| Cubukcu, Feryal | Dokuz Eylul University | Turkey |
| El Kashlan, Ahmed | Academy for Science and Technology | Egypt |
| Erickson, Jonathan C. | Washington and Lee University | USA |
| Fakher, Mahmoud | Ministry of Communication and Information Technology | Egypt |
| Kim, E-Jae | LG Electronics Inc. | South Korea |
| Ma, Yide | Lanzhou University | China |
| Malik, Om | University of Calgary | Canada |
| Mathew, Reejo | MYMIC- LLC | USA |
| Mourelatos, Zissimos | Oakland University | USA |
| Orsitto, Fulvio | University of Connecticut | USA |
| Prata, Alcina | Universidade Autónoma de Lisboa | Portugal |
| Smith, Lorenzo | Oakland University | USA |
| Sokolowski, John | Virginia Modeling Analysis and Simulation Center | USA |
| Wang, Xia | Oakland university | USA |
| Whiteley, Rick | Calabash Educational Software | Canada |
| Zamora, Inmaculada | Universidad del País Vasco | Spain |
| Zhang, Xiaoping | Lanzhou University | China |


## Number of Papers Included in these Proceedings per Country

(The country of the first author was the one taken into account for these statistics)

| Country | # Papers | % |
|---|---|---|
| TOTAL | 93 | 100% |
| United States | 24 | 25,81% |
| Japan | 11 | 11,83% |
| France | 8 | 8,60% |
| Italy | 7 | 7,53% |
| Germany | 5 | 5,38% |
| China | 3 | 3,23% |
| Czech Republic | 3 | 3,23% |
| India | 3 | 3,23% |
| Canada | 2 | 2,15% |
| Oman | 2 | 2,15% |
| Poland | 2 | 2,15% |
| Rwanda | 2 | 2,15% |
| Spain | 2 | 2,15% |
| Taiwan | 2 | 2,15% |
| Brazil | 1 | 1,08% |
| Denmark | 1 | 1,08% |
| Egypt | 1 | 1,08% |
| Indonesia | 1 | 1,08% |
| Iran | 1 | 1,08% |
| Iraq | 1 | 1,08% |
| Ireland | 1 | 1,08% |
| Lithuania | 1 | 1,08% |
| Mexico | 1 | 1,08% |
| Norway | 1 | 1,08% |
| Russian Federation | 1 | 1,08% |
| Saudi Arabia | 1 | 1,08% |
| Slovenia | 1 | 1,08% |
| South Africa | 1 | 1,08% |
| Turkey | 1 | 1,08% |
| Uganda | 1 | 1,08% |
| United Kingdom | 1 | 1,08% |

# Foreword

Engineering activities are based on the development of new Knowledge (*Scientia*), new 'made things' (*Techné*), and/or new ways of working and doing (*Praxis*). *Scientia*, *Techné,* and *Praxis* are three important dimensions of a comprehensive conception of Engineering as a whole. Engineering, as *Scientia*, is mostly developed in academia; as *Techné,* is practiced in industry generating technological innovations; and as *Praxis,* is carried out in technical and non-technical organizations, supporting managerial activities and technical procedures, via methodical and methodological design and implementation. This is why Engineering provides one of the most solid academic and professional substrata for bridging among universities, industries and governments.

Publications and conferences related to Engineering are usually oriented to one of its three dimensions. While this is an adequate thing to do when disciplinary focus is sought, it does not represent Engineering as a whole and it misses the very important synergic relationships among the three kinds of engineering activities mentioned above. This is why a group of scholars, professionals, and consultants, in the field of engineering, considered the possibility of organizing a conference where presentations would not be reduced to one specific Engineering dimension, but would foster the participation of academics, practitioners, and managers in the three dimensions of Engineering, in the same conference, so they can synergistically interact with each other. A consequence of this purpose is the organization of *The 4th International Multi-Conference on Engineering and Technological Innovation: IMETI 2011*: IMETI 2011, where submissions were accepted for the presentation of:

- **New knowledge** (Engineering as *scientia*);
- **New products and services**, i.e. technological innovations (Engineering as *techné*);
- **New technical and managerial methods and methodologies** (Engineering as *praxis*);
- **New meta-engineering** (Engineering of Engineering activities) knowledge, innovations, and methodologies.

The 8th International Conference on Cybernetics and Information Technologies, Systems and Applications (CITSA 2011) and The 9th International Conference on Computing, Communications and Control Technologies (CCCT 2011) have been organized in the context of IMETI 2011, because both are mainly oriented to Engineering and Technology. Both of them are International Multi-Conferences organized with the purpose of providing a communicational forum to researchers, engineers, practitioners, developers, consultants, and end-users of computerized, communications, and/or control systems and technologies in the private and the public sectors. This multi-disciplinary forum provides the opportunity to share experience and knowledge by facilitating

discussions on current and future research and innovation. Participants can explore the implications of relationships between new developments and their applications to organizations and society at-large.

One of the primary objectives of CITSA 2011, CCCT 2011 and, in general, IMETI 2011 is to promote and encourage interdisciplinary cross-fertilization and knowledge communication. They encourage systemic thinking and practice, including the analogical thinking that characterizes the Systems Approach, which is, in most cases, the required path to logical thinking, scientific hypothesis formulation, and new design and innovation in engineering.

CITSA 2011 and CCCT 2011 are spin-offs from the International Conference on Information Systems, Analysis and Synthesis (ISAS), and the World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI) which are yearly events that have been held in the last 17 years as a forum for Information Systems researchers, practitioners, consultants, and users who have been interchanging ideas, research results, and innovations in the area of Information Systems. Analytical as well as synthetical thinking represent the conceptual and methodological infrastructures that support the papers presented in ISAS conferences. Synthetical thinking supported papers in the Information Systems area, as well as in its relationships (analogies, "epistemic things", "technical synthetical objects", hybrid systems, cross-fertilization, etc.) with other areas.
The Organizing Committees of IMETI/CITSA/CCCT 2011 invited authors to submit original works, analogy-based hypothesis, innovations, experience-based reflections and concepts, specific problems requiring solutions, case studies, and position papers that explore the relationships among the disciplines of computers, communications and control, and the social and industrial applications within these fields.

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

1. the 636 members of the three Program Committees from 63 countries;
2. the 402 additional reviewers, from 62 countries, for their **double-blind peer reviews**;
3. the 277 reviewers, from 51 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 1431 reviews made by 679 reviewers (who made at least one review) contributed to the quality achieved in IMETI 2011. This means an average of 6.75 reviews per submission (212 submissions were received). *Each registered author could 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 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 IMETI 2011, including CITSA 2011 and CCCT 2011, about 212 papers/abstracts were submitted. These pre-conference proceedings include about 88 papers, from 28 countries, that were accepted for presentation,. We 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 probably appear in a more polished and complete form in scientific journals.

This information about IMETI 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 |
|---|---|---|---|---|---|---|---|
| **WMSCI 2011** | **391** | **1350** | **2461** | **1.82** | **6.29** | **193** | **49.36%** |
| **IMETI 2011** | **212** | **679** | **1431** | **2.11** | **6.75** | **88** | **41.51%** |
| **IMSCI 2011** | **276** | **856** | **2104** | **2.46** | **7.62** | **124** | **44.93%** |
| **CISCI 2011** | **388** | **973** | **2359** | **2.42** | **6.08** | **173** | **44.59%** |
| **TOTAL** | **1267** | **3858** | **8355** | **2.17** | **6.59** | **578** | **45.62%** |

We are also grateful 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 Honorary President of WMSCI 2011 and its collocated conferences, 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. Special thanks to Dr. C. Dale Zinn for chairing CCCT 2011 Program Committee (PC) and for co-chairing IMETI 2011 PC, to Professor Hsing-Wei Chu for co-chairing the IMETI 2011 PC and being General Co-Chair of CCCT 2011; to Professor Michael Savoie for being Co-General Chair of CCCT 2011 and CITSA 2011; to Professor José Ferrer for chairing the CITSA 2011 Organizing Committee; to professors Andrés Tremante and Belkis Sánchez for co-chairing the IMETI 2011 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 to Professors 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, IMETI 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,
*IMETI 2011 General Chair*

**IMETI 2011**
**The 4th International Multi-Conference on Engineering and Technological Innovation**
**The 8th International Conference on Cybernetics and Information Technologies Systems and Applications: CITSA 2011**
**The 2nd International Symposium on Security and Information/Communication Technologies: SICT 2011**
**The SUMMER 9th International Conference on Computing, Communications and Control Technologies: CCCT 2011**
**The 3rd International Symposium on Optical Engineering and Photonic Technology: OEPT 2011**
**The 3rd International Symposium on Energy Engineering, Economics and Policy: EEEP 2011**
**The 3rd International Symposium on Engineering Education and Educational Technologies: EEET 2011**

# VOLUME II
# (Post-Conference Edition)

## CONTENTS

## Computing, Communications and Control Technologies



## Control Systems, Technologies and Applications



## Hybrid Systems, Dual Technologies, Cross-Fertilizing Research and Analogical Thinking



## Optical Engineering and Photonic Technology

## Energy Policy



## Energy Systems/Technologies



## Social-Political, Economic and Environmental Contexts of Energy Systems and Technologies



## Educational Technologies



## Engineering Education

# Interaction between scientists, engineers and managers

Nader Ghahramany

Department of physics, Shiraz University

Shiraz, Iran

## ABSTRACT:

In the last two centuries natural academic scientists, such as physicists…have been swimming deep down the ocean of being to seek fundamental laws that govern the universe in microscopic (particles, nuclei, atomic..) and macroscopic (cosmology, gravity, galaxies..) scales and teach them to their students. The technology and business is located at shore where the knowledge is converted into money via industrial sites equipped with modern technology. Engineers must be swimming not as deep as scientists, but deep enough to keep contact with both scientists and technology and industrial managers at shore.

The scientists are solving fundamental problems while the engineers are involved with creative applied engineering problem solving with slightly different nature. They both utilize and also create technology. The managers, on the other side are occupied with the problems of the implementation of these findings to the Health, Happiness, and Safety of the society. New problems arise when the contacts between these three groups are weaken or even lost due to either deeper penetration of scientists in micro or macro scale worlds, causing distance to engineers or lost of proper contacts between engineers and industrial managers. These new problems are more pronounced and felt in the developing countries compared to developed ones.

Engineering schools are supposed to train intellectual individuals accepting the responsibilities of proper functioning of technology whereas the scientists are supposed to train open minded individuals equipped with realistic picture of our external world. Is it possible to establish academic institutions in which individuals are trained to be both, open minded and intellectual and at the same time being alert to the need of societies?.

Creation of interdisciplinary programs between these three groups might be the best solution. In order to do that, new undergraduate and graduate integrated curriculum powerful enough to train the right persons for such task should be established in both engineering and science colleges. Of course the detailed program might be slightly local and cultural dependent and needs to be discussed.

## Introduction

The curiosity and continuous struggle of man toward a better understanding of the universe and upgrading the quality of life is

endless. The shortness of human mean lifetime and deep comprehensibility time of numerous physical laws and diversity of advanced technology as external factors and multiplicity of individual aptitudes and interests as internal factors resulted in the differentiation of science into many different disciplines. Each discipline in time divided into more specialized sub-disciplines. For example in physics there are hundreds of recognized PACS (physics and astronomy classification scheme) numbers identifying the fields and subfields of physics. In addition to that, human integrity is to be understood and preserved. The physical and psychological well being and multi-dimensional aspect of human should be seriously considered in strategic developmental transitions.

I shall focus only on three main disciplines namely, science, engineering and corresponding managements though this discussion could be generalized to other interacting disciplines as well.

Before 19$^{th}$ century science was primarily theoretical research work. In 19$^{th}$ century science gradually found applications. The movement of locomotives and the flight of man are examples of such applications called engineering. Many engineering disciplines were born from science. For example, electrical, mechanical, civil, nuclear,… engineering were born from physics. Chemical, petroleum, textile,…engineering were born from chemistry. Some engineering may have more than one parent. For example material engineering were born from solid state physics and chemistry. Therefore discipline synthesize may create new applied engineering field.

In 20$^{th}$ century we witnessed not only the fast progress in engineering but also the creation of new engineering fields such as nano, earthquake and landslide, human factor, environmental and cosmogenic, … engineering.

As science and engineering became more specialized, communication among them became more important and vital. Any communication failure or weakening may result in a disaster or catastrophe threatening human life security. For example in three miles island accident (1979) or in Chernobyl plant disaster (1986), the role of communication is noticeable. In such accidents the threat involves safety of not only hundreds of employees working on the site but also innocent population surrounding the site. Strong scientific management is required for proper and secure functioning of such integrated plants.

In this paper it is intended to define three interacting groups involved namely scientists, engineers and managers and their relations, possibility of creating new interdisciplinary programs to promote and train more skillful and effective personnel and upgrade the integrity of the workers and the security of human life.

## Science and scientist

Science is a continuous process of learning and teaching of fundamental laws of being in microscopic and macroscopic scales by either logical (mathematical theory) reasoning or try and error (experimental data) methods. Science in not concern about the results of findings as much is concern about the findings. Process of finding is not subjective and remains judgment free fact. Such process involves

multiplication, unification, integration, differentiation, projection in order to determine the functional dependencies in terms of the variables. In each discipline different functions and variables may be defined but the scientific method is the same. Science collects the previous findings, correct them and generate new ones by deeper penetration into the micro or macro ocean of being.

Scientist is an individual human being acting as an honest detector whose duty is to observe, detect and report as accurate as possible. Scientist is not concerned about the interpretation and implication of the findings and the reports by others as much is sensitive about the accuracy of the findings. Objectiveness is a significant qualification of scientific report. The deeper a scientist swim into the ocean of being, the more prominent and pioneer becomes. Such qualifications require high individual integrity and social security to be provided by training institutions and by governments.

The communication and direct contact of scientists with each other via published journals and conferences is an indispensible condition for a productive scientific work. This follows from the objectivity, universality, collectivity and cumulative nature of science.

## Engineering and Engineer

Engineering is a continuous process of understanding the scientific findings and simultaneous analysis toward demanded applications in both micro and macro scales. In other words, engineering is an applied science. The main goal in engineering is to interpret and apply the previously and new scientific findings toward the demanded needs of public via technology. Such applications result in upgrading the quality of life.

An engineer is an individual that assumes the task of creative implementation of science. Engineer is not expected to swim as deep as scientist but deep enough to keep contact with them in order to be informed about the details and depth of involvement of scientific findings at the stage of application.

The process of application needs industrialization and technological investment. Therefore engineers need to keep contact with industrial managers and be aware of their potential and limitations. In some societies the scientific development is limited by low material standards due to rough correlation between economic gross national product and scientific size. This is because science and engineers need material support to flourish.

In addition, there are other non-material factors that play significant roles of limitation such as, lack of tradition, no public understanding of science and engineering and hostile bureaucracy. Nations with courage to risk and bold exploration of unknowns deep in ocean of being with highly motivated scientists and engineers posses great potential for industrial and technological development.

## Industrial and engineering management

Public and business is located at shore and is not aware of science and engineering exploration and research. Here, scientific findings and engineering designs are converted into high technology and are sold to public. Therefore huge factories and

industrial sites operate to manufacture and distribute the products designed by engineers who implemented findings of scientists.

In addition to technical challenges, engineering managers need soft skills such as business, communication, presentation and leadership skills to be effective. They are expected to assume responsibility for profits and losses to be recognized as strategic leaders. Therefore they must be trained by specialized institutions.

## Communication Dependency

The proper functioning of these three groups namely, scientists, engineers and managers depend upon sufficient communication and interactions. Scientists need to contact engineers as much engineers need to be in touch with industrial and business managers. Such communication becomes essential in strategic programs for a realistic and meaningful development. New problems such as individual alienation, uselessness of research outcome and lack of support arise when communication between these groups is weaken or lost due to either deeper penetration of scientists in the ocean of fundamentals causing distance between scientists and engineers swimming above them or lost of contact between engineers and managers located at shore, informing public about scientific findings and engineering creativities via introducing high technology products. Training is also required for such communication maintenance skills.

## Proposal

Engineering schools are set up to train intellectual individuals to apply and interpret the fundamental laws of nature and accept the responsibilities of proper functioning of the produced technology to serve public interests. Science colleges are supposed to train open minded individuals capable of projecting a real and objective picture of the universe and be equipped with the previously found fundamental laws and skillful enough to seek new laws. The managers are not supposed to have the skills of scientists or engineers to be successful in their assumed mentioned responsibilities. Is it possible to establish academic institutions in which individuals are trained to be both, open minded and intellectual and at the same time being alert to the need of societies?.

Creation of interdisciplinary programs between these three groups might be the best solution. In order to do that, new undergraduate and graduate integrated curriculum powerful enough to train the right persons for such task should be established in both engineering and science colleges. Of course the detailed program might be slightly local and cultural dependent and needs to be discussed.

# AMERICAN DEPOSITARY: A CASE STUDY FOR BRAZILIAN MARKET

**André Machado Caldeira**
**Fuzzy Consultoria Ltda, Brasil**
Av. Nossa Senhora de Copacabana 1376/302-Copacabana-Rio de Janeiro-RJ
CEP 22070-010

**Giovanna Lamastra Pacheco**
**Faculdades IBMEC/RJ, Brasil**
Av. Presidente Wilson, 118 – CEP. 20030-020 - Centro - Rio de Janeiro – RJ
Telephone: 00 55 21 91894497 ; Fax: 00 55 21 45034040
e-mail: gpacheco@ibmecrj.br

**Walter Gassenferth**
**Faculdades IBMEC/RJ, Brasil**
Av. Presidente Wilson, 118 – CEP. 20030-020 - Centro - Rio de Janeiro – RJ
Telephone: 00 55 21 45034000 ; Fax: 00 55 21 45034040
e-mail: walterg@quanticac.com

**Maria Augusta Soares Machado**
**Faculdades IBMEC/RJ, Brasil**
Av. Presidente Wilson, 118 – CEP. 20030-020 - Centro - Rio de Janeiro – RJ
Telephone: 00 55 21 45034000 ; Fax: 00 55 21 45034040
e-mail: mmachado@ibmecrj.br

## ABSTRACT

*Specialists often question market efficiency. Some works suggest arbitrage opportunities in several financial operations. Such opportunities can be explained mainly by information asymmetry, since pricing in the stock market is directly linked to information; therefore, the investor that has access to such information the soonest has a competitive advantage. The objective of this paper is to verify the existence of arbitrage opportunities via ADRs, traded in the American market, and their respective stocks, which are traded in the domestic market. Through a case study conducted with four companies, not considering the transition costs, arbitrage opportunity windows were found. Among the companies studied, two had frequent arbitrage opportunities, for one of them the arbitrage opportunity can be shaped by the time series model.*

Key words: Arbitrage, Stocks, ADRs.

## INTRODUCTION

The lack of long-term financial sources in the Brazilian capital market has led to an increase in the number of Brazilian companies issuing ADRs (American Depositary Receipts) with an aim at obtaining international visibility, providing more liquidity to the negotiations of their bonds and taking international funds at rates lower than those of the domestic market. (Camargos, 2003)

Such increase in the amount of ADRs issuance tends to make the markets involved more efficient, since international information exchange is increased. Efficient markets are beneficial to the economy and are of great interest to both large investors, due to the reduction in analysis costs, and small investors, since these are placed in a favorable competitive position in face of the possible investment choices, for the information available is already reflected in the prices.

At the same time, since the ADRs are papers issued in the United States market but have linking liabilities in the stocks of companies traded in the domestic market, there are arbitrage possibilities between the operations and the stocks issued in the national capital market. For instance, such opportunities arise as a result of distinguished taxation for the foreign investor, reduced transaction costs, different working hours of the stock markets, exchange flotation, and different transparency standards on disclosing information and negotiation practices (Rodrigues, 1999).

This paper aims at looking into the arbitrage possibilities in

ADR purchase and sale operations by comparing the prices of stocks of Brazilian enterprises in the domestic market and the prices of ADRs in the North-American market. In the present case, we disregard taxation, brokerage fees and other charges and we focus solely on the differences between the prices.
This article is structured on the basis of this introduction; the second part of it addresses some important concepts on market efficiency and the ADRs and its subsequent section is a case study that verifies the arbitrage possibilities. Finally, the conclusions are addressed in the fourth section.

# CONCEPTS

## MARKETING EFFICIENCY

The Efficient Market Hypothesis (EMH) was proposed by Fama in 1965 in "O comportamento de Preços de mercado de Estoque" (Inventory Market Price Behavior). According to his 1970 paper, a market is said to be efficient when the prices fully reflect all the information available.

In the first decade after it was created, the Efficient Market Hypothesis became a theoretical and empirical success, ending up providing the theoretical basis for many researches in the financial area over the seventies and eighties. At first, these were dedicated to predicting prices on the basis of historical data; now, in the eighties, such information compounds as dividends (Fama, 1988) and profit/price ratio became the basis for such outlook (Campbell and Shiller, 1988).

The level at which the prices are affected, given the information available, characterizes the type of efficiency of these markets. The literature distinguishes three levels of efficiency (Fama, 1970):

Poor efficiency: stock prices fully reflect the information contained in the historical price movements. Since price movements are completely independent from historical movement, then it is impossible to reach higher profits on the basis of past knowledge.

Semi-strong efficiency: the market is efficient in a semi-strong sense if the prices respond instantaneously and correctly to recently published information. This implies that there is no advantage in analyzing information that is publicly available because as soon as the information becomes public, it is immediately incorporated to prices. This includes profits and dividend disclosure, Incorporation & Acquisitions, publication of rights and any other kind of information on the company's stocks.

Strong efficiency: in this case, the prices reflect not only the information disclosed, but also all of the relevant information, including private data. Not even one of the company's employees would be able to obtain abnormal profits due to his position because the information is rapidly absorbed by the market.

These three efficiency levels lie upon three basic assumptions:

- The investigators are supposedly rational. They estimate the Net Present Values of their cash flows by using their respective discount rate. When the new information on the company becomes public, there is a rapid response that incorporates all of the information available almost immediately.

- Even if there are some irrational investors, their businesses are casual and, therefore, prices are not affected. According to the market efficiency hypothesis, since the steps taken by these irrational investors are not correlated, they end up by offsetting themselves.

- The third argument to the market efficiency hypothesis is based upon the notion of arbitrage (Friedman, 1953) arguing that in the market there are rational arbitrators that eliminate the influence of irrational investors on prices.

Since the irrational investors are buying stocks above their fair value and selling stocks below their fair value, they profit less than passive investors or arbitrators. In comparison to other investors in the market, these earn less and, as Friedman, 1953, points out, they cannot lose money infinitely because they would become illiquid or, eventually, disappear from the market. In the long run, if the arbitrage does not eliminate the influence of irrational investors in the market prices, the market forces will.

## AMERICAN DEPOSITARY RECEIPTS

In order to take advantage of the growing interest and allow for ease of access of investors to the capital of national companies and, this way, increase foreign capital attracting, the major Brazilian enterprises make use of such instruments as the ADRs.

In Brazil, according to a definition by the Central Bank, American Depositary Receipts are certificates representing stocks or other securities, representing rights and stocks, issued abroad by an institution called "Depositary", which have substance on securities issued by Brazilian companies deposited in specific custody in Brazil.

In a general sense, ADRs are papers issued and traded in the capital market in the United States with substance on stocks of non-American companies.

In Brazil, the ADR was created by Administrative Act 1927, of May 18th 1992, to stimulate the stock market. Some of the companies to issue ADRs were: Ambev, Aracruz, Bradesco, Brasil Telecom, Cemig, Companhia Brasileira de Distribuição (Pão de Açúcar), Copel, Embratel, Embraer, Gerdau, Itaú, Petrobras, Vale do Rio Doce, Companhia Siderúrgica Nacional, Telebras, Telesp Celular, Telemig e Unibanco.

The issuance of ADRs brings advantages both to the company and to investors. For the latter, it allows for a participation in a foreign market, and for the former, it provides for a wider

international presence with a consequent liquidity increase of its stocks, in addition to a new source of financing at a low cost.

Only those investors with a foreign bank account (individual or corporate) can buy and sell ADRs. Opening the bank account is a legal procedure, as long as it is stated and in compliance with taxation rules. Remittance of money abroad is only possible through financial institutions that are accredited to exchange transactions and, if the overall amount exceeds US$10 thousand, the Central Bank must be informed.

Investors can also convert their ADRs into company stocks and trade them in the company's country of origin. It is called arbitrage operation that operation in which the investor sees distortions between the price of the same asset in two different trading environments and takes advantage of it. For example: if the price of the stocks of a company (after the application of the conversion factor into ADR and then into dollars) is US$ 20 at Bovespa and US$ 21 in the United Stated, there is an opportunity to buy here, make the conversion and sell there at the same time, thus keeping the difference. The conversions are made through the custodian bank. The buyer here informs the bank that they wish to make a conversion and immediately sell the paper there. In the US, the Bank of New York is the main ADRs custodian institution and, in Brazil, Banco Itaú, for instance, holds a large part of custodies.

There are three levels of ADR, each with growing demands of transparency and adequacy to the American standards. When traded over-the-counter, the ADRs need not follow the rules of the Sarbanes-Oxley Act. This is a North-American Act that is also applicable to non-American companies that have stocks listed in the United States stock market (NYSE, AMEX and Nasdaq). It imposes standards of corporate governance such as the certification of financial statements by the CEO - chief executive officer – and by the CFO - chief financial officer – of companies.

## ADR LEVEL I

The ADR - Level I has the lowest level of demands and is traded in the American over-the-counter (OTC) market. The ADRs Level I provide their issuers a simple and efficient means of forming a group of investors with few legal requirements and mandatory reports. They are traded in the North-American OTC market and also in some stock exchanges outside the United States. Establishing an ADRs Level I program is considered to be the first step forward into the American stock market.

ADRs Level I Characteristics:

- Traded in the over-the-counter market (outside the organized market);
- There cannot be public offer in the United States;
- Need not comply with American accounting standards;
- It is no issuance of new stocks;
- Stocks bought in the secondary market;
- Its plain objective is to place stocks in the American market, preparing the ground for future primary stock issue;
- It is the simplest method, for it needs not meet all of the demands by USA's Securities Exchange Commission – SEC.

## ADR LEVEL II

The ADRs - Level II are traded in the stock exchanges of the USA (for example, the Nasdaq). It should be stressed out the in levels I and II there is no new issuance of new stocks. This level requires more statements and reports than in Level I. The company is bound to forward its accounting statements in US GAAP, United States Generally Accepted Accounting Principles: A set of accounting rules, conventions, standards and procedures used to produce financial information according to the models established by the FASB - Financial Accounting Standards Board, to the SEC, besides other reports. The company cannot raise funds by issuing ADRs of this level and adherence to this level does not provide an IPO. Generally, the companies that issue ADRs in this level are already taking part in the American stock market with ADRs Level I.

ADRs Level II characteristics:

- There cannot be public offer in the United States;
- Financial statements must be according to US GAAP;
- Must meet American accounting standards;
- It is no issuance of new stocks;
- More requirements from the SEC because ADR registry in Stock Exchange is mandatory;
- They are traded in the Stock Exchange.

## ADR Level III

The ADR Level III is traded in a national-wide stock exchange in the United States or in the Nasdaq, bound to a public offer in the United States of the stocks deposited. It is issued on the basis of new stocks issued by the company. The financial statements must be according to the US GAAP - United States Generally Accepted Accounting Principles. Therefore, the ADR – Level III has the same degree of demand as the ADR - Level II, but there is fund raising, since it has a substance on new stocks.

ADRs Level III characteristics:

- More complete and onerous;
- Must meet the requirements of the SEC and the Stock Exchange;

• Must comply with the American accounting standards;

• Its objective is to raise funds for the company;

• Full compliance with the SEC requirements;

• Like Level II, it must foresee a top quality institutional publicity.

## STANDARD 144-A

Also standing out is the ADR ruled by Standard 144-A, which authorizes a company to trade its stocks with communities of qualified institutional investors in order to streamline the liquidity of the American private issuance market. Besides the partial submission to the SEC requirements, this standard represents advantages because it is the most economic, fast and easy-to-raise-funds form, in addition to the negotiation through Nasdaq's Portal (private offerings, resales and trading through automated linkage). However, it is an inaccessible form to the stocks that have already been registered in American Stock Exchange.

# CASE STUDY

The data used in this paper refer to the period between April/1/2003 and November/7/2005, corresponding to a total of 618 observations per day for each asset and its respective ADR. The assets used were from Petrobras (PETR4), Telemar (TNLP4), Vale do Rio Doce (VALE5) and Banco Itaú (ITAU4).

The data used can be divided into three variables:

• Daily closing price of the ADRs of Brazilian enterprises traded in the North-American market in dollars (effective price);

• Daily closing price of Brazilian stocks, corresponding to the ADRs, traded in Real in the Brazilian market;

• Daily exchange rate, Real (R$) per dollar (US$).

The data related with the ADRs were obtained from the Economática, whereas the data on Brazilian stocks were obtained from the webpage of EasyInvest, a broker, and, finally, those data regarding the daily exchange rate were obtained from the webpage of IPEADATA.

These companies were chosen due to the level of importance of these assets in the BOVESPA index. From all the companies in the sample the dates that did not represent negotiations in both markets were excluded because there was no business with either the stock or the ADR. This fact is due to the difference of the dates of Brazilian and American holidays.

A new variable, called "ADR theoretical price", was calculated on the basis of the closing prices of the company's stocks, the dollar rate and the number of corresponding stocks to each ADR. That is to say:

$$PT_{it} = \frac{PA_{it} * N}{C_t}$$

where:

PTit = ADRi theoretical price of (US$) at instant t.

PAit = Stocki Price (R$) at instant t.

N = Number of corresponding stocks for each ADR.

Ct = Daily exchange rate (R$/US$) at instant t.

This calculation allows for a comparison between the "ADR effective price" (PE) and the ADR theoretical price" (PT), since the currencies and amounts were equaled. By this, a comparison test of unconditional means (Test T-Student for two pair samples for the means) can be performed. By disregarding the transition cost, the unconditional means of these two variables are expected to converge to the same point. This result is expected, since should the means be differing, there is room for continuous arbitrage.

The software used in this study was SPSS – Statistical Package for the Social Sciences, version 10.0. In the first moment, the level of linear dependence (Pearson Correlation) between variables PE and PT for each of the companies in the sample was studied. As expected, the calculated correlation coefficients show high linear dependence between the series; all companies had correlation above 99%. (Table 1).

**Table 1: Correlations of Test T for two pair samples for the means**

| Company | Observations | Correlation | Sig. |
|---|---|---|---|
| Petrobras | 618 | 99,967% | 0,000 |
| Telemar | 618 | 99,733% | 0,000 |
| Vale | 618 | 99,973% | 0,000 |
| Itaú | 618 | 99,971% | 0,000 |

Despite the high correlation, it cannot be said that there are no arbitrage possibilities between these markets. It can be concluded that if such opportunity does exist, it is not so common for a chance to appear.

After this verification, an option was made on working with a single variable for each company, which is expressed through the ratio between PE and PT. That is:

$$R_{it} = \frac{PE_{it}}{PT_{it}}$$

where:

PEit = ADRi Effective Price at instant t

PTit = ADRi theoretical price at instant t

With this variable, it is possible to verify the per cent difference between the theoretical price and the effective of the ADR of the i-th company.

In case of an efficient market (disregarding the transaction costs), this variable should equal to one, which is not verified. Therefore, when this variable is greater than one, there is an opportunity window, which can be used by acquiring stocks at the local market and selling the ADRs at the American market. When this ratio is lower than one, arbitrage is given as of the purchase of the ADRs at the US market and the sales corresponding to this ADR at the Brazilian market. In order to statistically test whether the unconditional mean of this ratio is equal to one, the T-Student test was applied for one sample (Table 2). This test points to frequent arbitrage possibilities for companies Telemar and Vale do Rio Doce, since the T test rejected (5% level) the hypothesis that the mean for such ratio is equal to one. As for companies Petrobras and Itaú, it cannot be said the same, once the hypothesis tested can be accepted (5% level).

**Table 2: T-Test for a sample (One-Sample)**

| | Test Value = 1 | | | | | |
|---|---|---|---|---|---|---|
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
| | | | | | Lower | Upper |
| PETR4 | -1,693 | 617 | ,091 | -4,888E-04 | -1,06E-03 | 7,818E-05 |
| TNLP4 | -2,790 | 617 | ,005 | -8,421E-04 | -1,43E-03 | -2,49E-04 |
| VALE5 | -3,702 | 617 | ,000 | -1,162E-03 | -1,78E-03 | -5,46E-04 |
| ITAU4 | -,295 | 617 | ,768 | -8,536E-05 | -6,54E-04 | 4,829E-04 |

As this test is used to test the unconditional mean of the data series, it only manages to indicate those assets with frequent arbitrage opportunities. However, the arbitrage possibilities may appear in some windows, which are not constant. In order to verify that, Table 3 shows the descriptive statistics for the series related with the ratio variable for all companies used.

It can be observed (Table 3) that the series of this variable for Petrobras has a minimum of 0.96, pointing to a 4% distance between the theoretical and effective prices of the ADR; in spite of having an unconditional mean statistically equal to one, this instant can present arbitrage possibilities, that is, there may be arbitrage windows. The same can be said of company Itaú, but the difference between maximum prices reached 3%.

**Table 3: Descriptive Statistics**

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| PETR4 | 618 | ,96 | 1,03 | ,9995 | 7,177E-03 |
| TNLP4 | 618 | ,93 | 1,03 | ,9992 | 7,504E-03 |
| VALE5 | 618 | ,93 | 1,02 | ,9988 | 7,802E-03 |
| ITAU4 | 618 | ,98 | 1,03 | ,9999 | 7,194E-03 |
| Valid N (listwise | 618 | | | | |

In an attempt to forecast the opening of arbitrage windows, the autocorrelation (ACF) and the partial autocorrelation (PACF) of the series were studied in order to verify whether the conditional means for the series are constant. Moreover, the Ljung-Box Test was applied with an aim at verifying whether the autocorrelation is statistically significant.

The variable studied regarding company Petrobras has significant ACF and PACF (Figure 1 and Figure 2, respectively) for the first lag. For companies Telemar and Itaú there is no statistically significant linear dependence for any lag (Figure 3, Figure 4, Figure 7 and Figure 8). The series for the variable related with company Vale do Rio Doce also has significant ACF and PACF (Figure 5 and Figure 6), which indicates a certain predictability by using the ARIMA models. The Ljung-Box test indicated (by using a 5% level of significance) that the series of Petrobras and Vale do Rio Doce have a significant linear dependence; by this, these data can be modeled by ARIMA models.

**Figure 1: Autocorrelation PETR4**

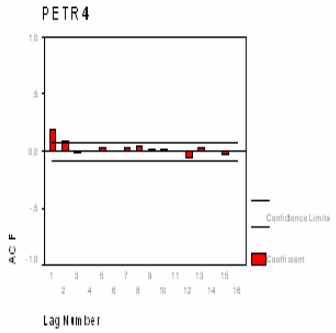



**Figure 2: Partial Autocorrelation PETR4**

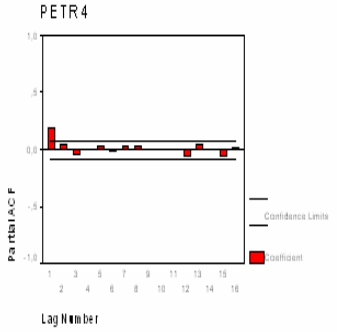



**Figure 3: Autocorrelation TNLP4**

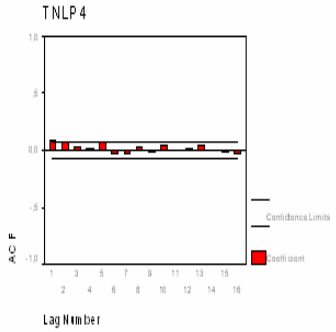



**Figure 4: Partial Autocorrelation TNLP4**

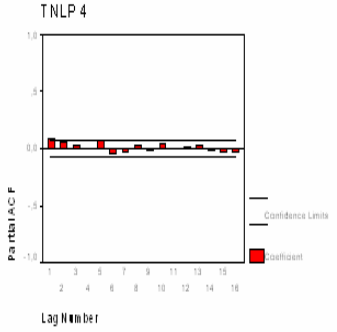



**Figure 5: Autocorrelation VALE5**

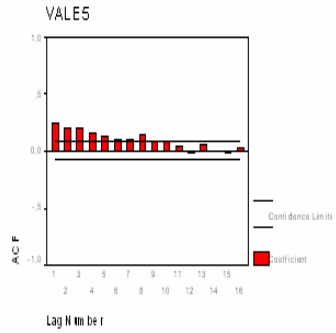



**Figure 6: Partial Autocorrelation VALE5**

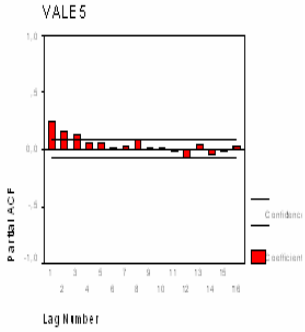



**Figure 7: Autocorrelation ITAU4**

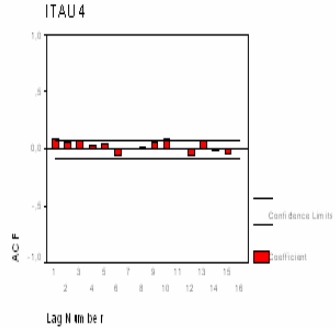
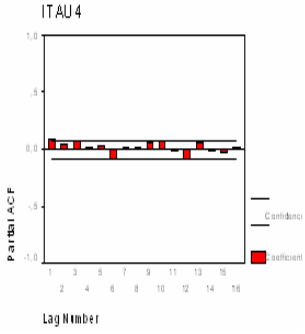



**Figure 8: Partial Autocorrelation ITAU4**



The ARIMA model that is best adjusted for the Petrobras series is AR (1), the estimated coefficient of which is 0.2858 and the constant is 0.9988. As for the Vale do Rio Doce series it is the ARMA (1,1) with self-regressive coefficient (AR) of 0.9510, movable mean coefficient (MA) of 0.7555 and constant of 0.9980.

These models allow for predicting the result of the ratio between the effective price and the theoretical price of the ADRs, and by this, it helps make the decision as to when to perform the operation and when to render resources available in order to put it into effect. Once the opportunity windows are modeled, the investor will be able to foresee when the market will adjust itself by closing the arbitrage window.

# CLOSING CONSIDERATIONS

The case study reflects the information asymmetry existing between the Brazilian and American markets. Such information asymmetry, in many cases, is corrected through arbitrage, that is, investors who are alert to the opportunity windows modify supply and demand by adjusting the price to the fair price.

The T-Student test for the means detected that in two of the four companies tested the ratio between the theoretical price and the effective price is different from one. This result suggests that there are frequent arbitrage opportunities using the ADRs of these enterprises. Now, for the other two companies, a 3% and 4% difference was detected, indicating that, despite the arbitrage opportunities not being frequent, there may appear some opportunity windows.

By analyzing the linear time dependence of the series through the ACF and PACF graphs and the Ljung-Box test, the possibility of predicting the ratio between the effective and theoretical prices of the ADRs was verified in two of the companies, and, by this, ARIMA models were estimated to model the time behavior of that series; such results may be used to decide the time to perform the operation, thus minimizing such risks as the liquidity of the papers.

This paper introduces an academic study which, due to some simplifications it cannot be applied directly, requires some changes. One limitation is not using the transaction rates, which may render the arbitrage operations unfeasible. Therefore, a more complex study comprising all transaction rates can be designed in the future aimed at both practical and academic effects.

# REFERENCES

CAMARGOS, M. A., GOMES, G. D. e Barbosa, F. V. (2003). *Integração de Mercados e Arbitragem com Títulos Transfronteiriços: ADRS – American Depositary Receipts*. **Caderno de Pesquisas em Administração**, São Paulo, 10, 51-67.

CAMPBELL, J. Y., and SHILLER, R. (1988). *Stock prices, earnings and expected dividends*. **Journal of Finance**, 43, 661-676.

FAMA, E. F. (1991). *Efficient capital markets II*. **The Journal of Finance**, Chicago: American Finance Association, 46, 1575-1617.

_____and FRENCH, Kenneth R. (1988). *Permanent and temporary components of*

*stock prices*. **Journal of Political Economy**, 96, 246-73.

_____(1970). *Efficient capital markets: a review of theory and empirical work*. **The Journal of Finance**, 25, 383-417.

FRIEDMAN, M. (1953). *The methodology of positive economics*, **Essays in Positive Economics**. Chicago: University of Chicago Press, 3-43.

RODRIGUES, E. L. (1999). *Segmentação, fragmentação e composição de ordens no mercado de capitais brasileiro: os efeitos da listagem de ações de empresas brasileiras no mercado norte americano através do mecanismo de recibos de depósito de ações*. **Tese (Doutorado em Administração)** – Instituto de Pós-Graduação e Pesquisa em Administração. Rio de Janeiro: UFRJ.

# A New Technique for Automatic Detection and Parameters Estimation of Pavement Crack

*Ghada Moussa, Ph.D.*
*Civil Engineering Department, Assiut University, Assiut 71515, Egypt*
Ghada.moussa@gmail.com

*Khaled Hussain, Ph.D.*
*Faculty of Computers and Information, Assiut University, Assiut 71515, Egypt*
Khaled.hussain2000@gmail.com

## ABSTRACT

Pavement condition evaluation is a significant part of a good pavement management system for effective maintenance, rehabilitation, and reconstruction decision-making. One of the key components of pavement condition evaluation is the quantification of pavement distresses data. Cracking is the main form of early pavement distresses. Cracking of pavement affects road condition, driving comfort, traffic safety, and consequently reduce pavement service life. Once initiated, cracking increases in extent and severity and accordingly accelerates the rate of pavement deterioration. Therefore, the awareness about crack type, extent, and severity is essential to evaluate pavement condition and to determine timing and cost of pavement maintenance. Digital image-based automated pavement evaluation has been gradually replacing the manual pavement evaluation due to its improved efficiency and safely operating.

In this paper, we are presenting a novel reliable automated pavement assessment system based on image processing techniques and machine learning methods. The proposed system has the ability to i) identify crack, ii) extract crack parameters, and iii) report the type, extent, and severity level of that crack in an output file. Actual pavement images were used to verify the performance of the proposed system. The results clearly demonstrated that the proposed system was able to automatically and effectively identify crack type and efficiently extract crack parameters from pavement images. Such information can be used by public road agencies to define maintenance plans and assist in pavement management decision-making, in accordance with real pavement condition.

**Key words**: Crack Detection, Flexible Pavement, Image Processing, Support Vector Machine

## 1. INTRODUCTION

Pavements are important infrastructures, they begin their life in excellent conditions and remain in excellent conditions for a few years without need of any maintenance. Over time, however, pavements exhibits distresses due to their constant usage and consequently their condition will worsen. Pavement distresses; visible undesirable imperfections on pavement surface, that affect pavement structural capacity, appearance and hence serviceability, are usually due to traffic loads, environmental conditions, and normal wear [1]. Pavement distresses represent a significant economic concern in any country. It is estimated that pavement distresses cause damage costing $10 billion each year in the United States alone [2]. Technically, cracks are the main form of early pavement diseases [3]. Unfortunately, if these early distresses were not treated, potholes are formed causing the pavement to become more dangerous. Therefore, a better and timely evaluation of pavement condition is likely to lower maintenance cost, support planning schemes and effectively allocating resources and increase drivers' safety and comfort [4]. To successfully conduct an adequate evaluation process, fast and reliable quantification of pavement distresses data using automatic systems are desired, instead of relying solely on the more conventional, time-consuming, labor-intensive and subjective, manual inspection procedures [5 and 6].

With the fast developments in computer technology, digital image acquisition, and image processing, many researchers have paid a great attention to use digital image-processing to produce automatic systems to assess pavement distress [7]. However, different types of distresses, complex texture and color of the pavement surface present a challenge in developing a precise yet reliable automated system for detection and evaluation of pavement distresses [8, 9, and 10].

To overcome the limitation of image-based automated systems, a novel reliable automated pavement assessment system based on image processing techniques and machine learning methods is proposed in this paper. Subsequently, the performance of the developed system is assessed and compared with ground truth data.

The rest of the paper is organized as follows. Section 2 briefly presents previous work. The various steps of the proposed system are explained in section 3. Section 4 presents the results of various experiments done to confirm the performance of the proposed system. Finally, section 5 draws some conclusions and presents some hints for future work.

## 2. PREVIOUS WORK

In the literature, different digital image-processing techniques such as, fuzzy set theory [11], Markov methods [12 &13], artificial life [14], neural networks [15, 16, and 17], and many other techniques have been used for crack detection and classification.

Cheng, et al. (1999), proposed a novel pavement cracking detection algorithm based on fuzzy set theory [11]. The main idea of their method is based on the assumption that the cracking pixels are always darker than their surroundings.

Delagnes and Barba (1995), proposed a Markov random field model for crack detection and extraction [12]. Another multi-scale approach for crack detection, using Markov random field mode, is presented by Chambone et al., 2009 [13].

An artificial living system for crack detection is proposed by Zhang and Wang (2004) [14]. In their study, a bottom-up approach was used for searching artificial structures within the pavement image. Bright points in images were excluded and the contrast between crack and non-crack pixels was enhanced. Qualitative experimental results were given in their analysis.

The neural network technique for pavement distress detection and classification was extensively adapted. Chou et al., (1994), used moment invariant and neural network for pavement crack classification [15]. After calculating moment invariants from different types of cracks thus obtaining crack features, neural network was used to classify these features. They trained their neural network and reported a one-hundred percent classification accuracy results. Cheng et al., (2001), used the mean and standard deviation as parameters and trained a neural network to select a threshold for pavement image segmentation [16]. Lee and Lee (2004), present an integrated neural network-based crack imaging system for crack type classification from pavement images [17]. They presented three neural; image-based, histogram-based, and proximity-based networks. Their classification was based on the sub-images (distress region) rather than crack pixels in digital pavement images.

Moreover, based on a spatial autocorrelation function, Lee and Oshima (1994) proposed an automated imaging procedure for crack identification and density measurement [18]. By calculating the autocorrelation function of a pavement image, the crack pattern in a noisy background could be identified. They concluded that their approach can identify crack type and density with a reasonable accuracy.

Based on local binary pattern operator, Hu and Zhao (2010) proposed a pavement crack detection approach [10]. They concluded that thin cracks (less than 1 mm) and cracks in strong texture were correctly and efficiently detected.

Based on image processing and pattern recognition techniques, Oliveira and Correia (2009) proposed an automatic system for crack detection and classification in flexible pavement images [19]. Their system was evaluated using real pavement images with their manually ground truth data. They reported promising results in both crack detection and classification.

Based on both image processing and supervised classification techniques, Younes et al. (2009) introduced an algorithm to classify pavement deterioration images [20]. An evaluation was conducted using 80 real pavement images, achieving 80% success in pavement deteriorations classification.

## 3. SYSTEM DESCRIPTIONS

A novel reliable approach for automatic crack detection, classification, and parameter estimation from flexible pavement images acquired during road surveys, based on image processing and machine learning techniques is presented. The proposed approach consists of four main stages: (1) segmentation, (2) feature extraction, (3) classification, and (4) parameters quantification.

### 3.1. Segmentation

Our goal is to accurately segment a given image into "crack" and "background" regions. We used the Graph Cut segmentation technique [21 &22] because it gives the best balance of boundary and region properties. The solution to our segmentation problem is a binary vector $A = (A_1, \dots, A_r, \dots, A_R)$, where $A_r$ equals one for crack and equals zero for background. To improve efficiency, we deal with these regions instead of image pixels. We used watershed algorithm [23] which divides the image into small regions ($R$).

We create a graph $G = (V, E)$ with nodes corresponding to regions $r \in R$ of the image. There are two additional nodes: a "crack" terminal (a source $S$) and a "background" terminal (a sink $T$). So, $V$ becomes

$$V = R \cup \{S, T\} \quad (1)$$

Each region $r$ has two t-links (terminal links) $\{r, S\}$ and $\{r, T\}$ connecting it to each terminal. *n-links* (neighborhood links) connect each pair of neighboring regions {p, q} in N. Therefore,

$$E = N \cup_{r \in R} \{\{r, S\}, \{r, T\}\} \quad (2)$$

We divide the regions into three types: crack, background, and unknown regions. One of the main characteristic of cracks is that it has a low intensity. Thus, the crack regions are identified by its mean intensity is less than a threshold $T_1$. In general, the intensity of the background is greater than the intensity of the cracks. Therefore, the background regions are identified by its mean intensity is greater than a threshold $T_2$. The K-means method is used to divide the crack and background regions into clusters. The mean intensity of the crack and background clusters are denoted as $\{K_n^C\}$ and $\{K_m^B\}$ respectively. For each region $r$, we compute the minimum distance from its mean intensity $M$(r) to crack clusters as $d_r^C = \min_n \|M(r) - K_N^C\|$. Then we compute the minimum distance from the region $r$'s mean intensity $M(r)$ to the background clusters as $d_r^B = \min_n \|M(r) - K_N^B\|$. The following table gives weights of edges in E.

Table1: Weights of edges in E

| Edge | Weight ($W_e$) | For |
|---|---|---|
| $\{r, S\}$ | $\infty$ | $r \in$ Crack Regions |
| | 0 | $r \in$ Background Regions |
| | $\frac{d_r^B}{d_r^C + d_r^B}$ | $r \in$ Unkown Regions |
| $\{r, T\}$ | $\infty$ | $r \in$ Background Regions |
| | 0 | $r \in$ Crack Regions |
| | $\frac{d_r^C}{d_r^C + d_r^B}$ | $r \in$ Unkown Regions |
| $\{p, q\}$ | $\frac{\lvert A_p - A_q \rvert}{1 + \lvert M(p) - M(q) \rvert}$ | $\{p, q\} \in N$ |

After defining the graph G, we need to calculate the minimum cost cut on the graph G. The cost of the cut is defined as $CH = \sum_{e \in H} W_e$. The minimum cost of the cut $CH$ can be calculated in polynomial time using the maxflow algorithm in [24].

After calculating the cut $H$, the segmentation binary vector $A = (A_1, \dots, A_r, \dots, A_R)$ is calculated as follows:

$$A_r(H) = \begin{cases} 1 \text{ (crack)} & \text{if } \{r, T\} \in H \\ 0 \text{ (background)} & \text{if } \{r, S\} \in H \end{cases} \quad (3)$$

### 3.2. Feature Extraction

One of the most difficult problems in the design of any computer vision system is the selection of a set of appropriate numerical attributes or features to be extracted from the region of interest for classification purposes. The success of any practical system depends critically upon this decision. Although there is little in the way of a general theory to guide in the selection of features for an arbitrary problem [25], it is possible to state some desirable attributes of features for identification of cracks; the features should be invariant with translation, scale, and light conditions.

Let $L$ be a two dimension binary array of size $m$ x n corresponding to the segmentation binary vector $A$. *Let*

$$(x) = \frac{\sum_{y=1}^{m} L(x,y)}{m}, for\ x = 1,2, \dots n\ , \text{ and} \quad (4)$$

$$SR(y) = \frac{\sum_{x=1}^{n} L(x,y)}{n}, for\ y = 1,2, \dots m. \quad (5)$$

Seven features are extracted as follows:

1. *PksNC* is the number of local peaks in *SC* that are greater than $T_3$ and that are separated by minimum distance of $T_4$.
2. *PksNR* is the number of local peaks in *SR* that are greater than $T_3$ and that are separated by minimum distance of $T_4$.
3. *PksMeanDistC* is the mean of the distances between peaks in *SC* and it equals zero if *PksNC equals one.*
4. *PksMeanDistR* is the mean of the distances between peaks in *SR* and it equals zero if *PksNR equals one.*
5. *PksStdDistC* is the standard deviation of distances between peaks in *SC* and it equals zero if *PksNC equals one.*
6. *PksStdDistR* is the standard deviation of distances between peaks in *SR* and it equals zero if *PksNR equals one.*
7. *PksLocMeanR* is the average of the normalized locations of the peaks in *SR.*

### 3.3. Classification

In this work, the Support Vector Machine (SVM) was used for the classification because it can produce accurate and robust classification results. The SVM is the state-of-the-art among classification algorithms [26].

The SVM is shown to be suitable for solving problems with a small sample set, nonlinearity, high dimension, over-fitting and local minima, eventually achieving a better generalization performance [27]. The proposed method is constituted essentially of two main phases; training phase and classification phase. As the name indicates, during the training phase is learned to recognize a set of different cracks. In its classification phase, it outputs the crack type (Transverse cracking, Longitudinal cracking, Block Cracking, or Alligator Cracking).

The training of the SVM is carried out in the following sequential order:

1. Scale each feature to the range [0-1].
2. Construct the training set of ($F_i$; $y_i$); $i$=1,2,…, $L$ where $F_i$ is the feature vector after scaling and $y_i$=(1,0,0,0) for Transverse cracking, (0,1,0,0) for Longitudinal cracking, (0, 0, 1, 0) for Block Cracking, or (0,0,0,1) for Alligator Cracking.
3. The RBF kernel ($K(x,y) = e^{-\gamma\|x-y\|^2}$) is used.
4. Use cross-validation to find the best parameter γ.
5. Use the best parameter γ to train the whole training set.

### 3.4. Parameters Quantification

The automated assessment of the crack extent and severity based on crack parameters (lengths and widths) are a useful input to pavement condition evaluation [28]. Our proposed system automatically processes crack images to compute the crack length and the average crack width for extent and severity level identification (low, moderate, or high) as defined in the Distress Identification Manual [1].

The crack length in meter is calculated using the following equations:

For Longitudinal crack,

$$LSC(x) = \begin{cases} 1 & if\ SC(x) \geq T_5 \\ 0 & if\ Sc(x) < T_5 \end{cases},\ length = \frac{\mathrm{WR}}{\mathrm{W}} \sum_{x=1}^{n} LSC(x) \quad (6)$$

For Transverse crack,

$$LSR(y) = \begin{cases} 1 & if\ SR(y) \geq T_5 \\ 0 & if\ SR(y) < T_5 \end{cases},\ length = \frac{\mathrm{HR}}{\mathrm{H}} \sum_{y=1}^{m} LSR(y) \quad (7)$$

As explained in the manual, the average crack width is the mean crack opening width of the crack [1]. To calculate the average crack width, we first need to calibrate the camera. Then we fit line(s) along the crack points and calculate the average crack width along the line(s). Our method is based on random sample consensus (RANSAC) method [29] and thinning process as shown in the following steps:

1. Let *WR* and *HR* be the actual width and height, respectively, in meter of the pavement captured by the camera. Let *W* and *H* are the width and height, respectively, in pixels of the pavement image. Since, the height of the camera is fixed at *CH* meter, the camera is pointed downward, and the camera parameters are fixed. We can assume that the width *PW* and height *PH* of a pixel in meter are equal to WR/W and HR/H, respectively.
2. Apply thinning process [30] for reducing the crack regions in the image *A* to skeletal lines that preserves the extent and connectivity of the original crack region while deleting most of the original crack pixels. The details of the thinning process is explained as follows:
   a. The image is divided into two different subfields in a checkerboard pattern.
   b. In the first sub-iteration, if conditions $G_1$, $G_2$, and $G_3$ are all satisfied, then pixel *p* is removed.
   c. In the second sub-iteration, if conditions $G_1$, $G_2$, and $G_3'$ are all satisfied, then pixel *p* is removed.

   Let $x_1$, $x_2$… $x_8$ are the values of the eight neighbors of *p*, starting with the east neighbor and numbered in CCW order.
   Condition $G_1$ is satisfied if $X_H(p)$ =1
   Where, $X_H(p) = \sum_{i=1}^{4} b_i$ and

$$b_i = \begin{cases} 1 & if\ x_{2i-1} = 0\ and\ (x_{2i} = 1\ or\ x_{2i+1} = 1) \\ 0 & Otherwise \end{cases} \quad (8)$$

   Condition $G_2$ is satisfied if $2 \leq \min(n_1(p), n_2(p)) \leq 3$
   Where, $n_1(p) = \sum_{k=1}^{4} (x_{2k-1}\ or\ x_{2k})$ and

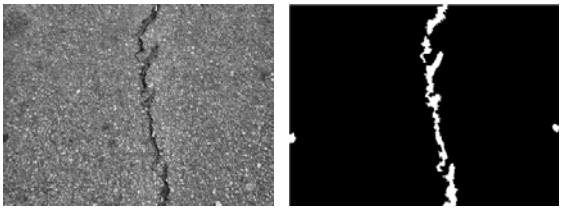
$$n_2(p) = \sum_{k=1}^{4} (x_{2k}\ or\ x_{2k+1})$$

   Condition $G_3$ is satisfied if $(x_2 or\ x_3 or\ \bar{x}_8)\ and\ x_1 = 0$
   Condition $G_3'$ is satisfied if $(x_6 or\ x_7 or\ \bar{x}_4)\ and\ x_5 = 0$
3. For fitting the line $y = mx+c$, we need to estimated the parameters *m* and *c* that minimize *E* where $E = \sum_i \frac{(mx_i - y_i + c)^2}{m^2+1}$
   a. Randomly select two points ($x_r$, $y_r$) and ($x_s$, $y_s$).
   b. Estimate the parameters *m* and *c* of the line that pass through the two points ($x_r$, $y_r$) and ($x_s$, $y_s$).
   c. For each crack pixel ($x_i$, $y_i$) caculate $d_i$ where $\frac{|mx_i - y_i + c|}{\sqrt{m^2+1}}$
   d. Count the number of outliers and inliers: the pixel ($x_i$, $y_i$) is consider outlier if $d_i$ is greater than a threshold τ; the pixel is consider inliers if $d_i$ is less than a threshold τ.
   e. Repeat steps a to d for *N* times and find the the parameters *m'* and *c'* of the line that have the maximum inliers.
4. The following steps show how we caculate the crack width, given the line $y = m'x+c'$.
   a. Set LineWidth =1
   b. Draw the line y = m'x+c' with width LineWidth in an empty image I1 of size W x H
   c. Set $\mathrm{D}_1 = |\mathrm{A} - \mathrm{I}_1|$
   d. Set $\mathrm{e}_1 = \sum\sum \mathrm{D}_1$
   e. Increase the LineWidth (LW) by one
   f. Repeat steps a to e M times and find the the line width LineWidth' (*LW'*) that has the minimam e1.
   g. Convert the line width from pixels to meter using the following equation:

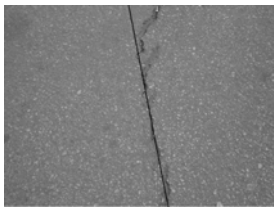
$$AverageCrackWidth = \sqrt{(LW' \cos(\tan^{-1}\left(\frac{-1}{\mathrm{m}}\right)))\frac{\mathrm{WR}}{\mathrm{W}})^2 + (LW' \sin(\tan^{-1}\left(\frac{-1}{\mathrm{m}}\right)))\frac{\mathrm{HR}}{\mathrm{H}})^2}$$

## 4. TESTING AND RESULTS

To test the reliability of the proposed system, a set of 87 pavement images (with 61 crack images) were collected from various road sections and used for the analysis. The implementation of our proposed method was carried out and the experimental results were recorded. Our system was trained to recognize two types of cracks: Longitudinal and Transverse. In Sahibsingh et al., it has been stated that "It is generally difficult to make a comparison of different recognition systems, even for the same problem, since different test sets are used for evaluating performance..." [25]. Therefore, no comparisons are offered for our proposed system with other systems. Examples of pavement images for Longitudinal and Transverse cracking and their testing results are shown in Figure 1 and 2 respectively. For both Figures a) represents original pavement image, b) represents crack image after segmentation, c) represents line fitting for avg. width calculations. From Figure 1 and 2, It can be observed that cracks are well detected using our system.

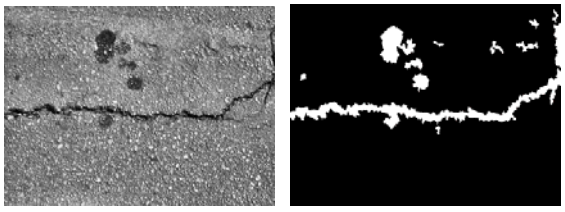


a) Original image b) Segmentation result



c) Line fit result

Fig. 1: Pavement surface image with a longitudinal crack



a) Original image b) Segmentation result

c) Line fit result

Fig. 2: Pavement surface image with a transverse crack

### 4.1. Crack Type Identification

In this section, results obtained when applying our system for classification of cracks depicted from real-life road images are discussed. The support vector machine was trained with 20% of the data set (17 images). Then we used this model to verify the other 80% (70 images). Our proposed system recognized the crack type of the test sample in less than 1 minute on a 64-bit Intel Core 2 Duo, 2.13 GHz, Ram 4GB, personal computer. The results of this test are summarized in Table 2. The system was able to detect the presence or absence of cracks in pavement images. Only 2 images out of 49 images with cracks were miss-detected. All the 21 distress-free pavement images were correctly predicted as no-crack. Therefore, the overall system prediction accuracy was found to be 97.1%. Based on these results, we can conclude that the proposed approach is able to detect pavement cracks effectively and accurately.

Table 2: Summary of system performance in crack type prediction

| Actual crack type | Total no. | System predicted crack type | | | Accuracy (%) |
|---|---|---|---|---|---|
| | | Long. | Transverse | No Crack | |
| Long. | 22 | 21 | | 1 | 95.5 |
| Transverse | 27 | | 26 | 1 | 96.3 |
| No distress | 21 | | | 21 | 100 |
| Total | 70 | 21 | 26 | 23 | 97.1 |

### 4.2. Crack Parameters Quantification

As the awareness about crack type is important, identifying the crack extent and severity level is essential to evaluate pavement condition and to determine timing and cost of pavement maintenance. The crack extent can be defined as the total cracked length for transverse and longitudinal cracking, as explained in the manual [1]. Moreover, the crack severity level is classified into three severity levels; low, moderate, and high based on the average crack width, as defined in the manual [1].
To evaluate the system performance in estimating average crack width and crack extent (length), field inspection data (actual measurements) was used. Paired t-tests were applied to compare the differences between the system estimated values and field inspection data, as shown in table 3 and 4. The statistical results indicate that there were no significant differences in the average crack width and crack extent between the system and field inspection data, at 95% confidence level (i.e. $\alpha = 0.05$). Note that the two miss-detected images were eliminated from the data set.

Table 3: Summary of statistic t-test for avg. crack width (mm)

| | N | Mean (mm) | S.D. | t | $t_\alpha$ |
|---|---|---|---|---|---|
| Actual Avg. Width | 59 | 8.90 | 6.12 | 0.7 | 2.05 |
| Predicted Avg. Width | 59 | 8.83 | 6.16 | | |

Table 4: Summary of statistic t-test for crack extent (m)

| | N | Mean (m) | S.D. | t | $t_\alpha$ |
|---|---|---|---|---|---|
| Actual Crack Extent | 59 | 1.190 | 0.514 | 0.25 | 2.05 |
| Predicted Crack Extent | 59 | 1.188 | 0.508 | | |

For the severity level prediction, table 5 presents severity-level comparison between in-the-field estimated values (actual severity level) and the system predictions ones. The results show that the system is able to predict the cracks severity levels in pavement images at in an average of 95.1% accuracy.

Table 5: Summary of system performance in severity prediction

| Actual Severity Level | No. of images | No. of images identified by the system | | | Accuracy (%) |
|---|---|---|---|---|---|
| | | Low | Moderate | High | |
| Low | 33 | 31 | 2 | | 93.9 |
| Moderate | 19 | 1 | 18 | | 94.7 |
| High | 7 | | | 7 | 100 |
| Total | 59 | 32 | 20 | 8 | 95.1 |

## 5. CONCLUSIONS

As the evaluation of pavement condition is the base of the pavement management, identification of pavement distresses and their parameters is an important step in this evaluation. Recently, the trend toward automatic distresses assessment has been widely spread, which is based largely on image processing techniques. However, the automatic recognition of pavement distresses from digital images is a difficult task, since pavement images can bring more information than those needed to detect a distress which can mislead the recognition process [20]. Based on image processing techniques and machine learning methods, we proposed a novel system for pavement distress assessment. Our proposed system was not only able to identify cracks but also classify cracks and quantify their parameters including extent and severity from digital pavement images, and present them in an output file. The proposed system attained promising performance results when tested via real pavement images with their ground truth data.

An area of future work that would be considered is increasing distress types for analysis. Furthermore, adapting this approach to process video streams would be studied.

## 6. REFERENCES


[1] J.S Miller and W.Y. Bellinger, **Distress Identification Manual for Long-Term Performance Program (Fourth Revised Edition)**, Office of Infrastructure Research and Development, Federal Highway Administration, Report No. FHWA-RD-03-031, 2003.

[2] R.J. Dilger, **American Transportation Policy**. West Virginia University, Department of Political Science. Westport, CT: Praeger Publishers, 2003.

[3] A. Ouyang, C. Luo and C. Zhou, "Surface Distresses Detection of Pavement Based on Digital Image Processing, **Advances in Information and Communication Technology**, Vol. 347, 2011, pp. 368-375.

[4] J. Kim, **Development of a Low-Cost Video Imaging System for Pavement Evaluation**, Ph.D. Thesis, Oregon State University, Oregon, United States, 1998.

[5] H.D. Cheng and M. Miyojim, "Automatic Pavement Distress Detection System", **Journal of Information Sciences**, ELSEVIER, Vol. 108, No.1, 1998, pp. 219-240.

[6] K.A. Abaza, S.A. Ashur, and I.A. Al-Khatib, "Integrated Pavement Management System with a Markovian Prediction Model," **Journal of Transp. Eng.**, Vol.130, No.1, 2004, pp. 24-33.

[7] K. McGhee, **Automated Pavement Distress Collection Techniques**, NCHRP Synthesis 334, National Cooperative Highway Research Program (NCHRP), USA., 2004.

[8] L. Li, P. Chan, and R.L. Lytton, "Detection of Thin Cracks on Noisy Pavement Images", **Transp. Res. Record**, No. 1311, pp.131-135, 1991.

[9] K. Wang, "Design and Implementations of Automated Systems for Pavement Surface Distress Survey", **Journal of Infrastructure Systems**, ASCE, Vol. 6, 2000.

[10] Y. Hu, C.X. Zhao, "A Local Binary Pattern Based Methods for Pavement Crack Detection", **Journal of Pattern Recognition Research,** Vol. 5, No. 1, 2010, pp. 140-147.

[11] H.D. Cheng, J.R. Chen, C. Glazier, C. and Y. G. Hu, "Novel Approach to Pavement Cracking Detection Based on Fuzzy Set Theory",**Journal of Computing in Civil Eng.**, Vol.13, No.4, 1999, pp. 270-280.

[12] P. Delagnes and D. Barba, "A markov random field for rectilinear structure extraction in pavement distress image analysis," **in Proceedings of IEEE International Conference on Image Processing (ICIP '95)**, Vol. 1, pp. 446–449, Washington, DC, USA, October 1995.

[13] S. Chambon, P. Subirats, and J. Dumoulin, "Introduction of a Wavelet Transform Based on 2D Matched Filter in a Markov Random Field for Fine Structure Extraction: Application on Road Crack Detection", **in IS&T/SPIE Electronic Imaging, Image Processing: Machine Vision Applications II**, San Jose, USA, 2009.

[14] H.G. Zhang and Q. Wang, "Use of Artificial Living System for Pavement Distress Survey", **Industrial Electronics Society, 30th Annual Conference of IEEE**, Pusan, Korea, Nov. 2004, pp. 2486-2490.

[15] J. Chou, W.A. O'Neill, and H.D. Cheng, "Pavement Distress Classification Using Neural Networks", **Systems, Man, Cybernetics1, IEEE**, 1994, pp.397–40.

[16] H.D. Cheng, J.L. Wang, Y.G. Hu, C. Glazier, H.J. Shi, and X.W. Chen, "Novel Approach to Pavement Cracking Detection Based on Neural Network", **Transp. Res. Record**, No. 1764, pp. 119-127, 2001.

[17] B.J. Lee, and H.D. Lee, "Position-invariant neural network for digital pavement crack analysis", **Computer-Aided Civil and Infrastructure Engineering**, vol. 19, no. 2, 2004, pp. 105–118.

[18] H. Lee, and H. Oshima, "New crack-imaging procedure using spatial autocorrelation function" **ASCE Journal of Transp. Eng.**, Vol. 120, No. 2, 1994, pp. 206-228.

[19] H. Oliveira, and P.L. Correia, **Supervised Crack Detection and Classification in Images of Road Pavement Flexible Surfaces,** Chapter in Recent Advances in Signal Processing, Austria, 2009.

[20] G. Younes, C. Hadda, and Z. Djellou, "Supervised Learning and Automatic Recognition of Asphalt Pavement Deteriorations", **MASAUM Journal of Basic and Applied Sciences**, Vol.1, No. 2, September 2009, pp. 254-259.

[21] Y. Boykov, and M. Jolly, "Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images", **in Proceedings of Eighth IEEE International Conference on Computer Vision (ICCV)**, Vol. 1, 2001, pp. 105-112.

[22] Y. Li, J. Sun, C.K. Tang, and H. Y. Shum, "Lazy Snapping", **SIGGRAPH04**, April 2004.

[23] L. Vincent, and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations". **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 1991, pp.583-598.

[24] Y. Boykov, and V. Kolmogorov, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision", **Energy Minimization Methods in Computer Vision & Pattern Recognition**, 2001.

[25] A.D. Sahibsingh, J.B. Kenneth, & B.M .Robert, "Aircraft Identification by Moment Invariants", **IEEE Transactions on Computers**, Vol. 26 No. 1, 1977, pp. 39-45.

[26] C.H. Lampert, "Kernel Methods in Computer Vision", **Foundations and Trends in Computer Graphics &Vision**, Vol.4, No.3, 2009, pp.193-285.

[27] V. Vapnik, **The Nature of Statistical Learning Theory**, New York: Springer-Verlag, 1995.

[28] A. Amarasiri, M. Gunaratne, and S. Sarkar, "Modeling of Crack Depths in Digital Images of Concrete Pavements using Optical Reflection Properties", **Journal of Transp. Eng.**, Vol. 136, No. 6, June 2010, pp. 489-499.

[29] M.A. Fischler, and R.C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting With Applications to Image Analysis and Automated Cartography”, **Communications of the ACM**, Vol. 24, No.6, 1981, pp. 381–395.

[30] L. Lam, S.W. Lee, and C.Y. Suen, "Thinning Methodologies-A Comprehensive Survey", **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol. 14, No. 9, 1992, pp. 869 - 885.

# Automatic Flow Analysis for Blood with Anticoagulant Using a Newly Developed Compact-Sized Falling Needle Rheometer

**Hideki Yamamoto, Takamasa Suzuki**
Department of Chemical, Energy and Environment Engineering, Faculty of Environmental and Urban Engineering, Kansai University, 3-3-35, Yamate, Suita, 564-8680, Osaka, Japan
e-mail: yhideki@kansai-u.ac.jp

**Kimito Kawamura**
Department of Process Engineering Technology, Research & Development Laboratories for Sustainable Value Creation, Asahi Breweries, Ltd, 1-1-21,midori, Moriya, 302-0106, Ibaraki, Japan

**Roberto Plasenzotti, Dominik Bernitzky**
Core Unit for Biomedical Research, Medical University of Vienna, Waehringer Guertel 18-20 / AKH1Q, A-1090 Wien, Austria

**Keywords: Flow property, Viscosity, Human blood, Rheometer, Flow curve**

**Abstract**

A compact-sized falling needle rheometer with quick operation and automatic flow analysis has been developed for viscometry of anticoagulated blood. The volume of a sample of blood only needs to be 4 mL and the measuring time is within 3 min. Measured flow properties of human blood and rabbit blood with anticoagulant are evaluated as a flow curve showing the relationship between the shear stress and shear rate.
The accuracy and the reproducibility of the presented rheometer are ascertained by viscosity measurements of standard liquid for calibration of viscometers manufactured by Nippon Grease Co., Ltd.. The standard liquids of JS10 and JS20 at 310.15K were chosen after careful consideration of the blood viscosity range (3.0 mPa•s to 10.0 mPa•s). Good uncertainty within 0.5 % and reproducibility within ±1.0% are confirmed by comparison with reference data of standard liquids.
Observed flow curves of the human and rabbit bloods with anticoagulant show three typical fluid regions, these are, the Non-newtonian fluid region for a low shear rate range of $0<\gamma>200$ $s^{-1}$, the transition region for the range about $150<\gamma<200$ $s^{-1}$ and the Newtonian fluid region for a high shear rate range of about 200 $s^{-1}<\gamma$. It is found that the range of blood apparent viscosities are 4.5 to 6.5 mPs•s for human blood and 3.5 to 4.5 mPs•s for rabbit blood. This paper is concerned with the flow analysis of fresh human blood viscosity without anticoagulant using a newly developed compact-sized falling needle rheometer.

## 1 Introduction

The rheological properties of the blood is not only one of the important factors in the pathological diagnosis of the human body, but also basic data essential for analytical study of the change of fluid mechanics of the blood arising from the deterioration of the health condition. The rheological properties of blood are also important factors in pathological nexuses, but basic data of the viscosity of anticoagulated blood, especially non anticoagulated blood, are still lacking. In this context, the development of viscometry with high accuracy and quick operation, as well as the establishment of a data evaluation method by pathology are largely required. However, currently, there is little observed data of blood viscosity from measurement immediately after collection of a blood sample in comparison with the viscosity data of blood added with anticoagulant. [1-4]
In research so far, it was found that the flow properties of human blood are available for preventive medicine for blood dyscrasia, clinical medicine, health care, functional foods, or the inspection of the effects of medicines. Also, it was reported that the viscosity of human blood influenced the concentration of fibrinogen in the plasma and the hematocrit value, and that the viscosity of human blood could offer important information for myocardial infarction and cerebral infarction, etc. However, the measurement of the viscosity just after blood collection from a body is not so easy, and the accumulation of numerical data of the viscosity is not yet sufficient. Most of the measurements of blood viscosity were carried out using blood added with anticoagulant, and at present there is little measurement of blood without anticoagulant. Also, it is problematical that many different viscometers are applied to these purposes, and a standard method has not yet been determined; hence, the values differ according to the type of device.
The difficulties of flow analysis of blood come not only from an aggregation-dispersion phenomenon of red corpuscle cells, but also its transformation property,

many interaction forces between corpuscle cells and blood plasma. Therefore, the establishment of exact viscometry for fresh blood is desired for further discussion about the relationship between blood disease and its flow properties.

In this paper, a compact-sized falling needle rheometer and a flow analysis method using this new device for fresh blood with anticoagulant have been developed, and the relationship between the apparent viscosity and physical properties of fresh blood has also been evaluated.

The theory of the presented viscometer is mainly based on the Stokes type of equation, and this is a kind of a falling body viscometer [5-9]. The viscosity of human blood can be measured with a small blood sample of about 3.5 $cm^3$ (total capacity is 4 $cm^3$) and with rapid operation within 2 min after taking a blood sample from the human body. The total scale of this compact-sized falling needle rheometer is downsized to about 1/30 the size of the previous apparatus [6]. A circular cylinder needle made of polypropylene is applied for the experiment, and its outer diameter and total length are 2 mm and 20 mm, respectively. This needle is also minimized at 1/5 that of a previous needle [6]. The density of the falling needle is controlled by the mass of a sinker enclosed in the needle tube. Flow analysis of the sample fluid is carried out using the needle's terminal velocity and the density difference between that of human blood and of a falling needle [10,11].

As stated above, the compact-sized falling needle rheometer is applied to the viscosity measurement of fresh blood before its coagulation. This paper is concerned with the development of a new rheometer for the measurement of flow properties of rabbit and human bloods.

**Fig. 1** Photograph of the compact-sized falling needle rheometer for measurement of fresh blood viscosity

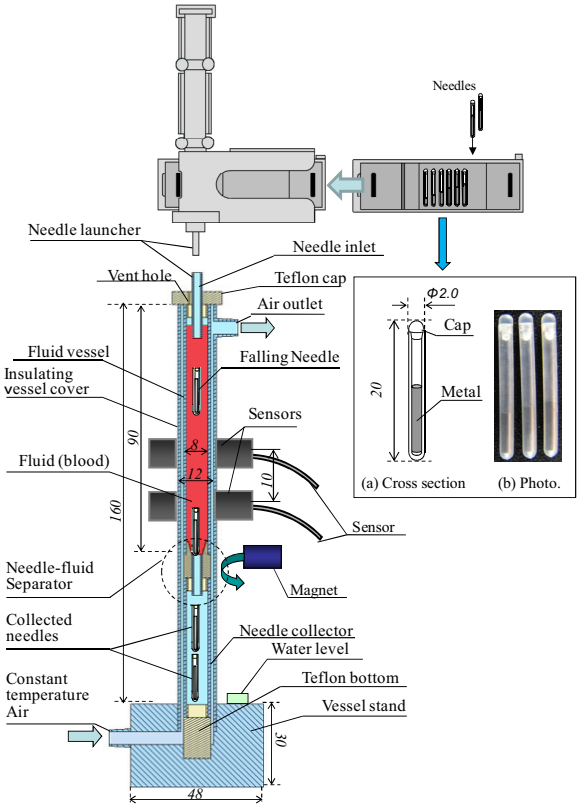



**Fig. 2** Schematic diagram of the compact-sized falling needle rheometer for measurement of human blood viscosity

## 2 Compact-Sized Falling Needle Rheometer with Automatic Operation

A schematic photograph of the compact-sized falling needle rheometer with automatic operation for measurement of blood viscosity is shown in Fig. 1. The schematic diagram of falling needle rheometer is illustrated in Fig. 2. The experimental apparatus consists of vertical double cylindrical vessels (one is a fluid vessel and the other is an insulating vessel cover) made of acrylic material (Fig. 2). The cap and bottom of the inner fluid vessel are made of Teflon. The inner fluid vessel for a blood sample is covered with the insulating vessel cover. The temperature of the inner fluid vessel is controlled at 310.15 K using a constant temperature air circulation by Peltier effect system within an uncertainty of 0.5 K. Constant temperature air is circulated in the space between the fluid vessel in the insulating cover. The diameter of the inner fluid vessel is 8.0 mm, and the height of the vessel is 90 mm. The total volume of the inner fluid vessel is about 4$cm^3$. A needle collector for the collection of the falling needles is connected to the bottom of the inner fluid vessel via a needle-fluid separator made of Teflon. The needle-fluid separator is a slender cylindrical tube, and its diameter is 2.2 mm, which is similar to the needle diameter (2 mm) shown in Fig. 2.

When a sample fluid (human or rabbit blood) is introduced into the fluid vessel, the sample fluid does not leak into the space of the needle collector because the pressure in the needle collector is controlled at atmospheric pressure. After the operation of the first needle dropping automatically using needle falling system, the needle stopped at the bottom of the fluid vessel shown in Fig. 2 and is rapidly manually moved into the space in the needle collector by the guidance of a small magnet from outside the vessel. This movement of the falling needle from the bottom of the fluid vessel to the needle collector is important and indispensable for rapid measurement of fresh blood within 2 min. It was found in the preparatory experiment that the leakage of sample fluid into the needle collector from the fluid vessel is very little.

As each of the parts of the experimental apparatus such as the fluid vessel, the needle collector, the insulating vessel cover and a vessel stand shown in Fig.2 can be taken apart easily, the collection of falling needles after the experiment can be very rapid and easy. This experimental apparatus is considerably compact for measurement of the viscosity of human blood compared with previous apparatuses.

**Table 1** Densities of falling needles used for viscometry of human blood and rabbit blood

| Needle No. | Density ($10^3$ kg·m$^{-3}$) |
|---|---|
| 1 | 1.090 |
| 2 | 1.118 |
| 3 | 1.212 |
| 4 | 1.278 |
| 5 | 1.360 |
| 6 | 1.400 |
| 7 | 1.438 |
| 8 | 1.563 |

The details of the falling needle used in this experiment is also given in Fig.2(a) and (b). This falling needle is a slender hollow cylindrical tube made of polypropylene. The diameter of the needle is 2 mm and its total length is 20 mm. The shape of both sides of each needle is hemispherical. Eight falling needles with different densities shown in Table 1 are used in this experiment. The density of each falling needle is controlled by the mass of a sinker (iron) enclosed inside the needle tube. This sinker is fixed at the bottom of the needle tube so that the center of gravity is at a lower position. The density is calculated by the volume and mass of each needle, and its uncertainty is estimated to be within ± $0.5\text{x}10^{-3}$ g・cm$^{-3}$. Table 1 shows measured needle densities used in the experiment, they are determined in consideration of average blood density (1.000g・cm$^{-3}$ to 1.100 g・cm$^{-3}$). .

In order to lead the falling needle to the center of the sample fluid, a needle inlet and needle launcher is equipped at the top of the fluid vessel. The needle launcher is a slender cylindrical tube as shown in Fig. 2. A pair of magnetic sensors is also installed at the middle part of the fluid vessel as shown in Fig. 2. The distance between magnetic sensors is 10 mm vertically. The passing time of each falling needle between magnetic sensors is automatically measured by a programmable controller manufactured by Sumitomo Metal Co., LTD. This magnetic sensor unit can be applied not only to clear liquids but also to opaque liquids. The programmable controller is connected to the personal computer via an amplification unit. It is possible to evaluate the falling velocity of each needle, and flow analysis such as a flow curve, apparent viscosity, and yield stress of the sample fluid can be measured automatically.

## 3 Experimental Method

Just after taking a blood sample from a vein, the blood (human blood or rabbit blood) is introduced into the fluid vessel shown in Fig. 2 and the top of the fluid vessel is covered with the Teflon cap in which the needle launcher is installed. This fluid vessel is equipped in the insulating vessel cover with two magnetic sensors. The vertical distance between the two magnetic sensors is 10 mm, and the fluid vessel is placed vertically on the vessel stand using a water level. Constant temperature air is circulated through the space between the fluid vessel and insulating vessel cover. The temperature of the fluid sample is kept at 310.15 K within an uncertainty of 0.5 K. Eight needles with different densities as shown in Table 1 are dropped down vertically in the sample fluid automatically using mechanical operation. The flow analysis is carried out using the observed passing time (terminal velocity) of the falling needles, needle densities, and blood density. Eight falling needles with different densities are used for measurement of the viscosity. Densities of blood are measured by the portable density/specific gravity meter (Kyoto Electronic Manufacturing Co., Ltd.) within an uncertainty of $10^{-4}$ g・cm$^{-3}$. The calibration of the presented falling needle viscometer is carried out using a standard liquid for calibrating a viscometers (JS5, JS10) manufactured by Nippon Grease Co., Ltd. EDTA-2Na manufactured by Wako Pure Chemicals Co. Ltd., is used as an anticoagulant for both bloods.

## 4 Fluid Analysis using Compact-Sized Falling Needle Rheometer

Figure 3a and 3b show the model for flow analysis and velocity distribution in the compact-sized falling needle rheometer. This model is based on the flow analysis around the falling circular cylinder (falling needle) in the static fluid introduced into the cylindrical vessel. In order to apply this model to the motion of a falling needle and the mass transfer, the following four conditions are assumed [12].

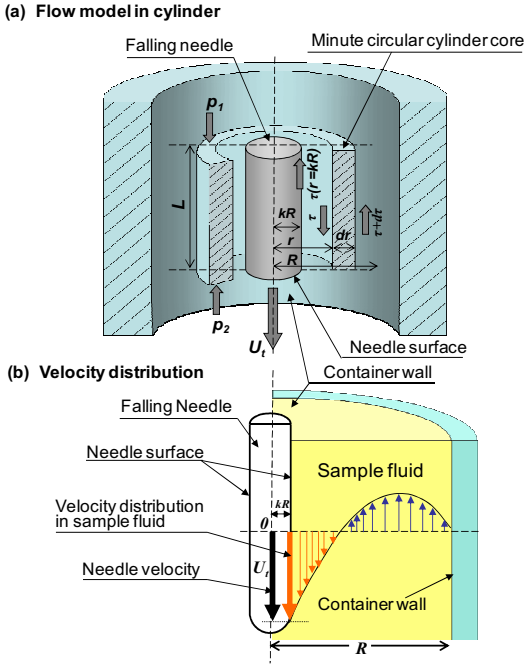



**Fig. 3** (a) Flow model and (b) velocity distributions of the sample fluid in the fluid vessel of falling needle rheometer

(1) Fluid is an incompressible liquid.
(2) Slipping between the falling needle surface and container wall is negligible.
(3) Flow in the cylindrical vessel is laminar flow.
(4) The falling needle free-falls in the center of the vessel with a terminal velocity.

These assumptions and the constitution equation of fluid are used for flow analysis. This flow model can be applied for many types of constitution equations shown in Fig. 4. In this work, the constitution equations of Newtonian and non-Newtonian fluids are applied for blood analysis. The flow model of a falling needle in a static fluid according to the above assumptions is given in Fig. 3a. This model shows that the falling needle falls at a terminal velocity ($U_t$) in the static fluid introduced into the cylindrical fluid vessel.

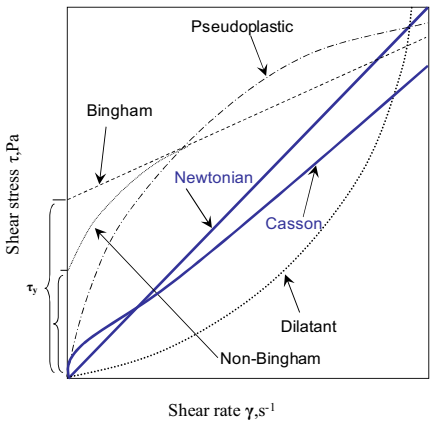



**Fig. 4** Typical flow curves for Newtonian and non-Newtonian fluids

The fluid vessel diameter is $R$, and $k$ is the ratio of the needle diameter to fluid vessel diameter.

The minute circular cylinder core shown in Fig. 3a is assumed as the fluid model for theoretical analysis. The inner diameter and outer diameter of this core are $r$ and $r$+d$r$, and the total length is L. The shear stresses on the inside and outside surfaces of the minute circular cylinder core are $\tau$ and $\tau$+d$\tau$, respectively. The pressures at the top and bottom of the minute circular cylinder core are $P_1$ and $P_2$. When the falling needle falls at a terminal velocity in the static sample fluid, the momentums affected on four surfaces of the minute circular cylinder core shown in Fig. 3a are balanced with each other, and they are balanced while the needle is falling at the terminal velocity. Therefore, this force balance can be described by the following equation:

$$\mathrm{P}_1\left\{(r+dr)^2\pi - r^2\pi\right\} + 2\pi rL\tau = \mathrm{P}_2\left\{(r+dr)^2\pi - r^2\pi\right\} + 2\pi(r+dr)L(\tau+d\tau) \quad (1)$$

When $\Delta P = P_1 - P_2$ is less than 0, Eq.1 is arranged as follows:

$$\frac{1}{r}\frac{\mathrm{d}(r\tau)}{\mathrm{d}r} = \frac{\Delta P}{L} \quad (2)$$

Furthermore, while the needle falls at the terminal velocity in the sample fluid, the force balance of gravity, buoyancy, pressure and shear stress affected on the needle surfaces are given as

$$(\rho_s - \rho_f)g\pi(kR)^2 L + \pi(kR)^2\Delta P = 2\pi kRL\tau_{(r=kR)} \quad (3)$$

In this equation, $\rho_f$ and $\rho_s$ are the fluid and needle density, respectively. The left-hand side first term of Eq.3 is the force of gravity and buoyancy, and the second term is the force of the pressure difference. The right-hand side term is the shear stress. This balance can be simply described by

$$(\rho_s - \rho_f)g + \frac{\Delta P}{L} = \frac{2\tau_{(r=kR)}}{kR} \quad (4)$$

Figure 3b illustrates the velocity distribution of the sample fluid due to falling of the needle. The amount of fluid (Q) to transfer between the falling needle surface and the container wall due to falling of the needle can be calculated by

$$Q = 2\pi\int_{kR}^{R} ur\,\mathrm{d}r = \pi(kR)^2 U_t \quad (5)$$

Figure 3b shows that the sample fluid around the falling needle is pulled downward with falling of the needle in the static sample fluid. On the other hand, the fluid near the container wall rises with the falling needle. The

maximum velocity in the sample fluid is that on the needle surface. The maximum velocity is equal to that of the falling needle velocity. On the other hand, the velocity on the container wall becomes zero according to the above assumptions. Therefore, the boundary conditions of the velocity distribution can be described by

$$u_{(r=kR)} = -U_t \tag{6a}$$

$$u_{(r=R)} = 0 \tag{6b}$$

In order to obtain the relationship between the shear rate and shear stress for the sample fluid, the Eqs. 2, 4, 5, 6a, and 6b and a constitution equation of the sample fluid are used simultaneously for flow analysis [6].

The constitution equation for a Newtonian fluid based on the law of viscosity is given by

$$\tau = \mu\left(\frac{\mathrm{d}u}{\mathrm{d}r}\right) = \mu\gamma \tag{7}$$

where $\mu$ is the viscosity, $\tau$ is the shear stress, and $\gamma$ is the shear state. The viscosity of the fluid sample can be calculated by the following equation from combining Eqs. 2, 4, 5, 6a, 6b, and 7.

$$\mu = -\frac{(\rho_s - \rho_f)g(kR)^2\left\{(k^2+1)\ln k + 1 - k^2\right\}}{2(k^2+1)U_t} \tag{8}$$

As $R$ and $k$ in Eq. 8 are fixed values according to the size of the fluid vessel and falling needle, they are arranged by the following equation using the geometric constant G:

$$G = -\frac{2(k^2+1)}{(kR)^2\left\{(k^2+1)\ln k + 1 - k^2\right\}} \tag{9}$$

Therefore, the viscosity of a fluid can be simply described by

$$\mu = \frac{(\rho_s - \rho_f)g}{GU_t} \tag{10}$$

The shear rate of the sample fluid on the falling needle surface can be obtained from Eqs. 7 and 9 as

$$\gamma_{(r=kR)} = \frac{\mathrm{d}u}{\mathrm{d}r} = \frac{(k^2-1)U_t}{kR\left\{(k^2+1)\ln k + 1 - k^2\right\}} \tag{11}$$

The shear rate distribution for the radius direction can also be calculated by this equation.The shear stress of the sample fluid on the falling needle surface can also be described by the following equation by substituting Eqs. 8 and 11 for Eq. 7:

$$\tau_{(r=kR)} = \frac{(\rho_s - \rho_f)g(1-k^2)kR}{2(k^2+1)} \tag{12}$$

Equations 11 and 12 become fundamental data for flow analysis of the sample fluid. In this experiment, eight needles with different densities are used for the flow analysis. A flow curve of the sample fluid is obtained from the relationship the between shear stress and shear rate for eight needles [6].

## 5 Results and Discussion

### 5.1 Accuracy of the Compact-Sized Falling Needle Rheometer

The accuracy and the reproducibility of the presented rheometer are ascertained by viscosity measurements of the standard liquid for calibration of viscometers manufactured by Nippon Grease Co., Ltd.. The standard liquids of JS10, and JS20 were chosen after careful consideration of the blood viscosity range (3.0 mPa•s to 7.0 mPa•s). Flow curves of standard liquid show in Fig. 5 and comparison of measured values with standard value are given in Table 2. Good uncertainty within 0.5 % and reproducibility within ±1.0 % are confirmed by comparison with reference data(Japnease Standard ; JS10, JS20), and this experimental apparatus is improved from previous work (previous uncertainty was 1.25 %) [6]. It is thought that this improvement of accuracy was caused by the manufacturing method of falling needles, that is, the presented needle is manufactured using a model made of metal with high accuracy.

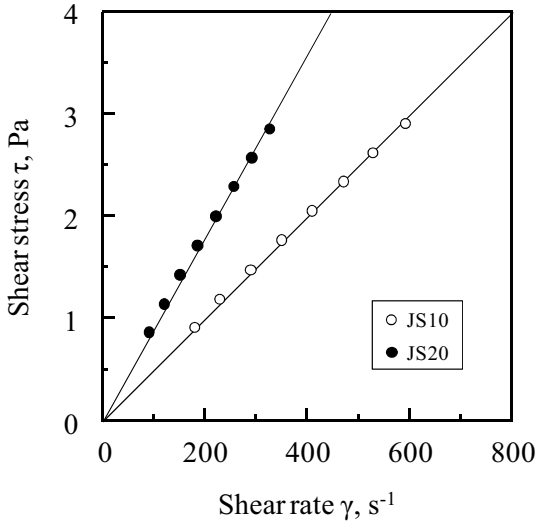



**Fig. 5** Flow curve of Standard liquids of JS10 and JS20 at 310.15K
JS-10, JS-20: Standard liquid for calibration of viscometers

**Table 2** Comparison of measured viscosity of standard liquids for calibration of vicometers with standard values at 310.15K

| Standard liquid | Viscosity[ mPa・s] | | Deviation[%] |
|---|---|---|---|
| | This work | Standard values | |
| JS 10 | 4.987 | 4.982 | 0.11 |
| JS 20 | 9.006 | 8.963 | 0.48 |

### 5.2 Flow Analysis of Human and Rabbit Bloods

Flow analysis of fresh human blood for female and male was carried out using a compact-sized falling needle rheometer using anticoagulant. Each result was evaluated as a flow curve, an apparent viscosity and hematocrit values. These rheological properties were compared between female and male blood.

In the experiment, fresh human brood of 20 $cm^3$ was taken from human veins, about 4$cm^3$ of whole blood was rapidly introduced into the fluid vessel without anticoagulant. The temperature of the fresh human blood was kept at 310.15 K using a constant temperature water bath. The anticoagulant (EDTA-2Na) was added to the other blood (17 $cm^3$), and that blood was kept at 278.15 K and sent to a medical center for the measurement of the hematocrit value.

At first, the first needle was introduced into the needle launcher shown in Fig. 2, and the passing time of the needle between two magnetic sensors was measured by a programmable controller. After this operation, the falling needle stopped at the bottom of the fluid vessel and was rapidly moved into the space in the needle collector by the guidance of a small magnet from outside the vessel through manual operation. Next, the second needle was also introduced into the needle launcher as soon as possible. When the final needle (eighth needles) was introduced into the needle launcher and the passing time was measured, the rheometric operation was finished. Densities of fresh human blood were measured by a portable density/gravity meter manufactured by Kyoto Electric Co., Ltd., within an uncertainty of $10^{-4}g \cdot m^{-3}$ as soon as possible. Human blood densities for females and males are listed in Table 3.

**Table 3** Apparent viscosity, hematocrit value, and density of human blood and rabbit blood at 310.15 K

| Sex | Ht[a] (%) | Density ($10^3$ kg·m$^{-3}$) | Fresh blood + anticoagulant (mPa·s) |
|---|---|---|---|
| Mail | 42.8 | 1.0531 | 5.96 |
| Female | 38.1 | 1.0511 | 4.96 |
| Rabbit I | 40.5 | 1.0402 | 3.65 |
| Rabbit II | 45.0 | 1.0541 | 4.84 |

[a]Hematocrit value: the volume percentage of red cells included in whole blood

The experimental treatment of the fresh human blood with anticoagulant was carried out within 120 s except for the time needed for taking the blood ( about 60 s), that is, the total time needed for this operation for each person was finished within 3 min.

The observed flow curve for male blood with anticoagulant was shown in Fig. 6. This flow curve showed a linear relationship between the shear stress and shear rate in a high shear stress range (150 $s^{-1}$ to 400 $s^{-1}$). However, non-Newtonian behavior (Casson behavior) was confirmed in a low shear stress range (0 to 150 $s^{-1}$).

The observed flow curve of fresh blood showed the three typical fluid regions, that is, the non-Newtonian fluid region for the low shear rate range, and the transition region and Newtonian fluid region for the high shear rate range. Figure 7 shows the flow curve of female human blood with anticoagulant and a similar tendency with Fig. 6 as non-Newtonian fluid was obtained.

This flow curve also shows a linear relationship between the shear stress and shear rate in the high shear stress range (130 $s^{-1}$ to 400 $s^{-1}$). However, non-Newtonian behavior (Casson behavior) was confirmed in the low shear stress range (0 to 130 $s^{-1}$).

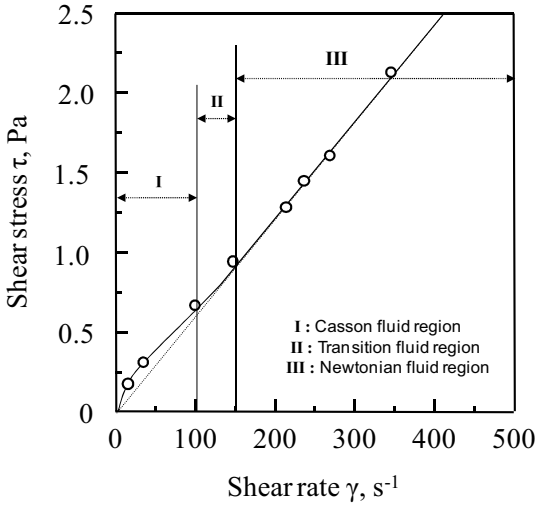



**Fig. 6** Flow curve of fresh human blood for male with anticoagulant at 310.15 K

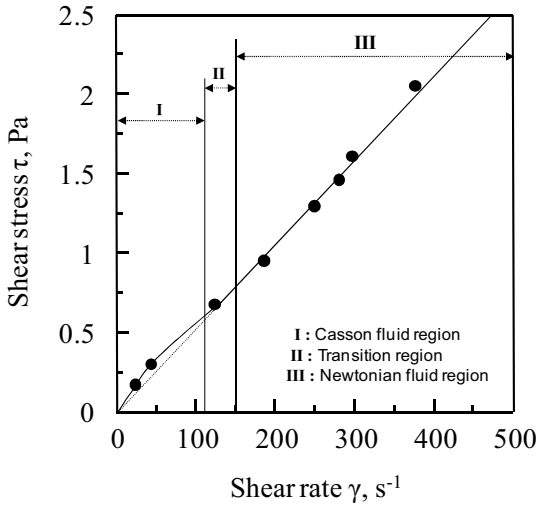



**Fig. 7** Flow curve of fresh human blood for female with anticoagulant at 310.15 K

Figure 8 is showing the flow curve of rabbit blood. Similar tendency for flow curve with human blood was obtained. Figure 9 gives the behavior of non-Newtonian regain of low shear ranges in the flow curve for rabbit blood without anticoagulant. This data was obtained from the experiments using special needles controlled similar density with rabbit blood density. It is found that this rheometer can measure flow curve for blood without anticoagulant.

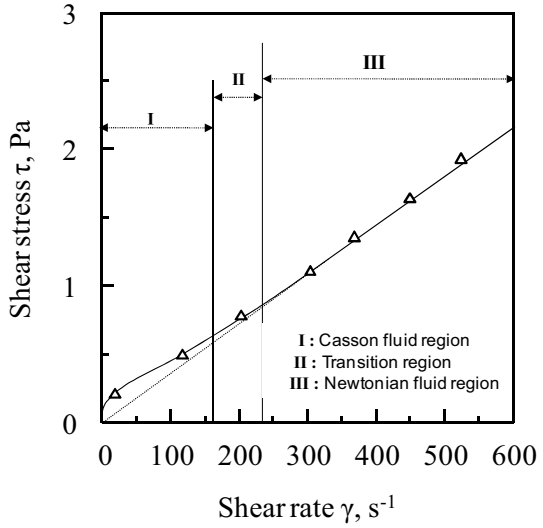



**Fig. 8** Flow curve of fresh rabbit blood with anticoagulant at 310.15 K

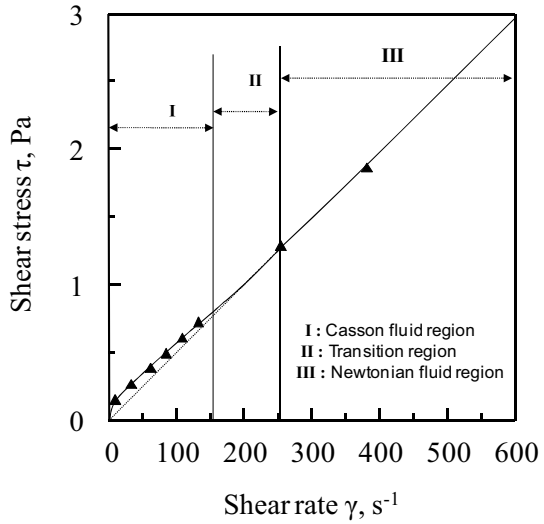



**Fig. 9** Flow curve of fresh rabbit blood without anticoagulant at 310 .15K

Figure 10 shows the relationship between the apparent viscosity and shear rate for fresh human blood and rabbit blood with anticoagulant at 310.15 K, and the rheological parameters that were obtained are listed in Table 3.
Figure 10 and Table 3 show a comparison between the flow properties of male and female blood. In the case of the apparent viscosity with anticoagulant, the blood viscosity of male blood was higher than that of female blood.

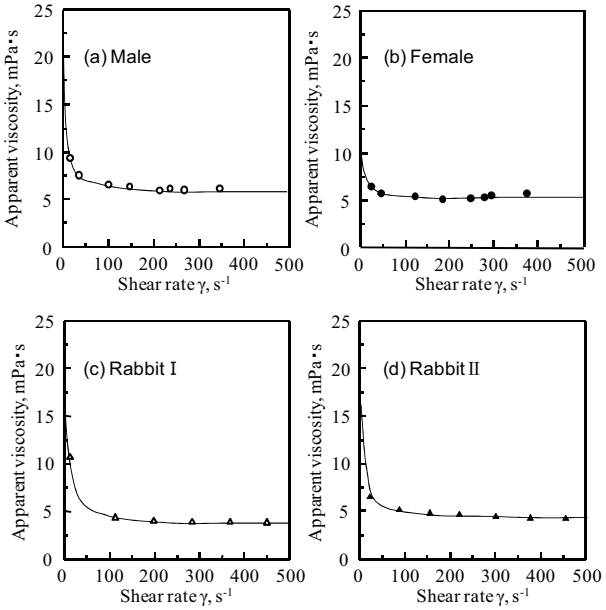



**Fig.10** Relationship between apparent viscosity and shear rate for fresh human bloods and rabbit bloods at 310 .15K

Figure 11 shows the relationship between the apparent viscosity for human blood for 21 person and hematocrit values that are the percentage of red cells included in whole blood.
A linear relationship with the hematocrit values was obtained. However, this linearity of apparent viscosity shows down with increasing of hematcrit value.
It was found that the viscosity for blood was closely connected with the hematocrit values.

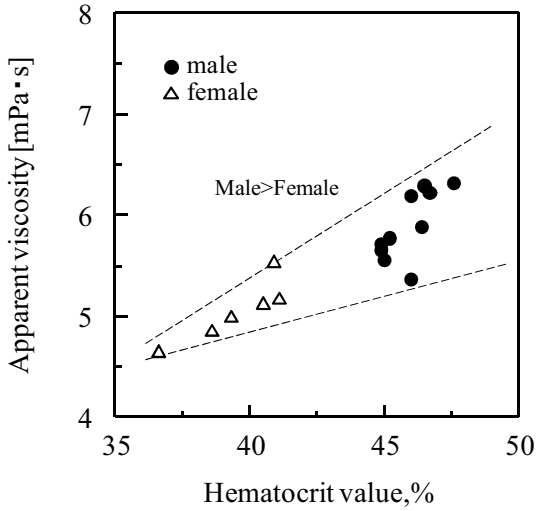



**Fig. 11** Relationship between average viscosity and hematocrit values of fresh human blood at 310.15 K

## 6 Conclusion

A compact-sized falling needle rheometer with quick operation and automatic flow analysis has been developed for the viscometry of human blood and rabbit blood with anticoagulant. The volume of a fresh blood sample only needs to be 4 $cm^3$ and the measuring time is within 3 min after taking a blood sample from the vein. The measured flow properties of both bloods are evaluated as a flow curve, that is, the relationship between the shear stress (τ) and shear rate (γ). Observed flow curves of fresh human blood show the three typical fluid regions, that is, the non-Newtonian fluid region for the low shear rate range, and the transition region and Newtonian fluid region for the high shear rate range. Flow properties of blood such as the apparent viscosity (μ) in the Newtonian fluid region are measured, and they are compared between male and female blood. It is found that the human blood viscosity of males (6.0 mPs･s to 6.4 mPs･s) shows a higher value than that of females (4.8 mPs･s to 5.3 mPs･s). A linear relationship between the hematocrit value, which is the volume percentage of red corpuscles in the human blood, and the apparent viscosity is observed for male and female blood. Rabbit blood viscosities with anticoagulant were also measured using this falling needle rheometer. It is found that the compact-sized falling needle rheometer presented in this work is very useful for rheometry studies of fresh human blood without anticoagulant.

## References


[1] E.W. Merrill, **Physiol. Rev**. 49, 863 (1969)

[2] M.F. Kiani ,A.G. Hudez, **Biorheology** 28, 65 (1991)

[3] H. Hartert, **Flow Properties of Blood and Other Biological Systems**, Pergamon Press, Oxford, New York & Paris, (1960),pp. 186-192

[4] P. Gaehtgens, **Biorheology** 17, 183 (1980)

[5] H.Yamamoto,K.Kawamura,K.Omura,S.Tokudome, **Int.J.Thermophysics**,31,2361(2010)

[6] H. Yamamoto, **J. Chem .Eng. of Japan** 25, 803 (1995)

[7] M. J.Davis, H. Brenner , **Phys. Fluids** 13, 3086 (2001)

[8] E.G. Wehbeh, T. J. Ui, R. G., **Phys. Fluids** 8, 645 (1993)

[9] J. A. Lescarboura, G. W. Swift, **AIChE J.** 14, 651 (1968)


## Nomenclature

| | |
|---|---|
| $d$ | needle diameter, m |
| $f_e$ | yield stress of a Casson fluid, Pa |
| $g$ | gravitational acceleration, $m \cdot s^{-2}$ |
| $G$ | geometric needle constant, $1 \cdot m^{-2}$ |
| $K$ | fluid consistency, $Pa \cdot s^{-n}$ |
| $k$ | ratio of container to needle diameter |
| $kR$ | needle radius, m |
| $L$ | total needle length, m |
| $n$ | fluid index |
| $P_1, P_2$ | pressure of the upper and lower end of aminute circular cylinder, Pa |
| $\Delta P$ | pressure difference ( $\Delta P = P_1 - P_2$), Pa |
| $Q$ | net flow rate of fluid pushed aside by the needle, $m^3 \cdot s^{-1}$ |
| $r$ | radius coordinate, m |
| $R$ | container radius, m |
| $u$ | velocity in the system length direction, $m \cdot s^{-1}$ |
| $U_t$ | terminal velocity of a falling needle, $m \cdot s^{-1}$ |
| $\gamma$ | shear rate, $s^{-1}$ |
| $\mu$ | Newton viscosity, Pa･s |
| $\pi$ | circular constant |
| $\rho_f$ | fluid density, $kg \cdot m^{-3}$ |
| $\rho_s$ | needle density, $kg \cdot m^{-3}$ |
| $\tau$ | shear stress, Pa |
| $\tau_y$ | yield stress, Pa |

# Needed Capabilities for Future Wireless Sensor Networks and Road Traffic Simulators

**Evarist NABAASA**
**Department of Networks, Makerere University**
**enabaasa@must.ac.ug**
**Kampala, +256, Uganda**

**and**

**Tonny E. BULEGA**
**Department of Networks, Makerere University**
**tbulega@cit.mak.ac.ug**
**Kampala, +256, Uganda**

## ABSTRACT

This paper provides an overview of the needed capabilities for current road traffic simulators and wireless sensor networks simulators intended to cooperate to solve common problems in Intelligent Transport Systems. Intelligent Transport Systems simulators have tended to be domain specific. In this work we present the needed capabilities of combining a simulator for wireless sensor networks and a simulator for road traffic scenarios. In future if the suggested capabilities are implemented within the simulators, researchers will quickly solve more pressing problems in Intelligent Transport Systems.

**Keywords:** Intelligent Transport Systems, Wireless Sensor Networks, Simulators.

## 1. INTRODUCTION

Simulation tools are emerging and have started assisting in understanding of Intelligent Transport Systems (ITS) models, algorithms, designs and requirements among others [1]. Current road traffic simulation tools [2] [3] [4] tend to specialize in simulation of road networks and the traffic that move on them.

Wireless Sensor Networks (WSNs) simulators [5] [6] also tend to specialize in simulation of WSN scenarios. The deep specialization of simulators has made them to be too domain specific, this has limited the gains one would see by combining them.

Some attempts to combine road traffic simulators with wireless sensor network simulators have been made [1] [7] with many unexplained assumptions.

We explore the capabilities that will enhance simulators to easily solve more compelling Intelligent Transport Systems problems. The rest of this paper is arranged as follows; in section II we discuss related work, section III we present wireless sensor network simulators, section IV we present road traffic simulators, section V we discuss the needed capabilities for combining simulators, and finally we conclude in section VI.

## 2. RELATED WORK

In [1] a Traffic Control Interface (TraCI) is used to interconnect Simulation of Urban Mobility (SUMO) with TinyOS SIMulator (TOSSIM) to have the two simulators exchange information, in this research there are many assumptions that were not clarified. In their work of Mobility Support for Wireless Sensor Networks Simulations for Road Intersection Safety Applications, it was assumed that the simulated vehicles running on a SUMO road network were

automatically carrying the sensor nodes to form a Car2Car sensor node communication but this is not practically supported by SUMO.

The authors in [7] built an Adaptive Traffic Light Control with Wireless Sensor Networks. In their work, they used Green Light District (GLD) simulator to simulate the road network with junctions, GLD is a microscopic road network simulator. They also used magnetometer sensors to form a sensor network for detecting vehicles at a junction. Their paper does not show how the magnetometer sensor network they used communicates with GLD simulator. The Green Light District simulator does not have an inbuilt support for simulating WSNs.

## 3. WIRELESS SENSOR NETWORK SIMULATORS

Wireless Sensor Networks have started gaining more ground on the stage of computer networks. Given the need of deploying many sensor nodes in a WSN, simulations become a more realistic choice in researching and understanding of their Application, Protocols, Design, Topologies, Security among others. These simulators help researchers to avoid the otherwise large and expensive real hardware network deployment. A considerable number of simulation tools have been developed. These tools tend to differ in many features, such as programming language, visual environment, supported protocols, freely available or with costs [1]. Below we give an overview of the common ones.

**Ns2**

Network Simulator (NS2) [6], is based on two languages: C++ for the simulator environment, and OTcl (Object oriented Tool Command Language) for executing user command scripts. Via OTcl scripts, the user is able to define particular network topologies, specific protocols, Applications he wishes to simulate and the form of output he wishes to obtain from the simulator. NS2 is a discrete event simulator, where the advance of time depends on the timing of events which are maintained by a scheduler. The Network Animator (NAM) application is used to graphically visualize the designed network and how the sensor nodes communicate.

**Omnet++**

OMNeT++ [8], is an extensible, modular, component-based C++ simulation library and framework, primarily for building network simulators. "Network" is meant in a broader sense that includes wired and wireless communication networks, on-chip networks, queuing networks, and so on. Domain-specific functionality such as support for sensor networks, wireless ad-hoc networks, Internet protocols, performance modeling, photonic networks, etc., is provided by model frameworks, developed as independent projects. OMNeT++ offers an Eclipse-based IDE, a graphical runtime environment, and a host of other tools.

**Tossim**

TinyOS Simulator (TOSSIM) [9], is a discrete event simulator for TinyOS sensor networks. Instead of compiling a TinyOS application for a mote, users can compile it into the TOSSIM framework, which runs on a PC. This allows users to debug, test, and analyze algorithms in a controlled and repeatable environment. TOSSIM replaces low-level components of a TinyOS application by simulation-specific code to simulate radio chip, Analogue-to-Digital Converter (ADC) components and timer hardware [1]. The extension TOSSIM-SF is a serial forwarder extension which allows communication from outside to single nodes while simulation is in progress. Under TinyOS the compiler can be forced to produce code for TOSSIM by just adding the option *sim* or *sim-sf* respectively to the compiler call. TOSSIM allows simulation of very large networks (thousands of nodes), network links are simulated at bit level and it also allows asymmetric links.

Other popular Wireless Sensor Network Simulation tools are JiST/SWANS [10] and OPNET [11].

## 4. ROAD TRAFFIC SIMULATORS

Simulation of road traffic usually caters for simulating, road networks, mobility of individual participants, rules followed by the participants and sometimes, rules for traffic lights control at intersections. Road traffic simulators support either of the following traffic models: macroscopic, mesoscopic and microscopic.

Macroscopic simulation models analyze traffic flow in which the description of reality is shifted from individuals to "averaged" variables like flow and density.

Mesoscopic simulation models use methods of statistical mechanics to provide a probability function which expresses the probability to have a certain vehicle at a certain position.

Microscopic simulation, simulates each entity (car, train, person) of reality individually by considering all relevant properties. A number of microscopic simulation environments have been developed [12]. Below is an overview of some common mobility simulators.

**Sumo**
Simulation of Urban Mobility (SUMO) [3], is an open source, highly portable, microscopic road traffic simulation package designed to handle large road networks. It can perform GUI or Command line based simulations, import various types of maps, consider multiple lanes, traffic signs and right-of-way rules and configure individual routes for vehicles [1]. It does not have inbuilt sensor network modules though it allows communication with such networks via TraCI.

SUMO, uses poison distribution to generate vehicles which are injected into the road network. XML scripts are written to generate nodes, links, routes and the network files among others, finally basing on the network file, a configuration file is generated for the simulation scenario.

Fig. 1. Simulated T-Junction

With a simulation scenario like in Fig. 1, SUMO provides inbuilt Traffic Light Control (TLC) models on how the traffic lights handle the vehicles that are in or out bound of the T-Junction. With a WSN for detecting vehicles along the T-Junction lanes, the TLC models supported by SUMO would achieve more meaningful results.

**Green Light District**
GLD allows us to edit infrastructures using the mouse, to set different spawning frequencies creating different traffic patterns, to compare a large number of implemented controllers (new ones can be easily added), and to evaluate the controllers using different statistical measures (such as average waiting time of cars). The simulator itself is based on a cellular automaton model [13], and therefore a microscopic model which can be used for modeling various amounts of detail forexample the road user could be a car, bus, police car.

**Infrastructures:** An infrastructure consists of roads and nodes. A road connects two nodes, and can have several lanes in each direction (see Fig. 2). The length of each road is expressed in units. A node is either a junction where traffic lights are operational (although when it connects only two roads, no traffic lights are used), or an edge-node.

**Agents:** There are two types of agents that occupy an infrastructure; vehicles and traffic lights. All agents act autonomously, following some simple rules, and get updated every time-

step. Vehicles enter the network at the edge-nodes.

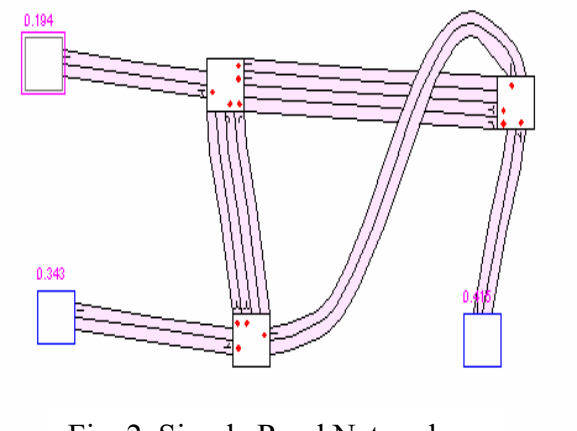



Fig. 2. Simple Road Network

Each edge-node has a certain probability of generating a vehicle at each time step. Each vehicle that is generated is assigned a destination, which is one of the other edge-nodes. The distribution of destinations for each edgenode can be adjusted. There are several types of vehicles, defined by their speed, length, and number of passengers. The state of each vehicle is updated every time step. It either moves with the distance given by its speed, or stops when there is another vehicle or a red traffic light ahead. At a junction, a car decides to which lane it should go next according to its driving policy. Once a car has entered a lane, it cannot switch lanes.

**Controllers:** Every junction is controlled by a traffic light controller (TLC) that decides on the best configuration of red and green lights. A TLC will only consider safe configurations, that is, configurations in which moving cars do not intersect. A TLC can share information with other controllers to improve global performance. GLD has several built in TLCs, and allows for custom TLCs. GLD lucks standardized interface for communicating with other outside applications.

Other common mobility simulators are Sparamics [15], VISSIM [14], JSim [16], among others.

## 5. NEEDED CAPABILITIES FOR COMBINING INTELLIGENT TRANSPORT SYSTEMS SIMULATORS

Currently mobility simulators and WSNs simulators are domain specific. Each of them tries to perfect modules in its domain without direct support of the other. For road traffic simulators to communicate with WSNs simulators, one has to newly hard-code a module or interface to enable data to be shared among them. The following are some of the needed capabilities that we feel future simulators should automatically support in order to realize smooth data sharing.

### Inbuilt Sensor Nodes

The road traffic simulators like SUMO among others should have and support inbuilt sensor nodes.

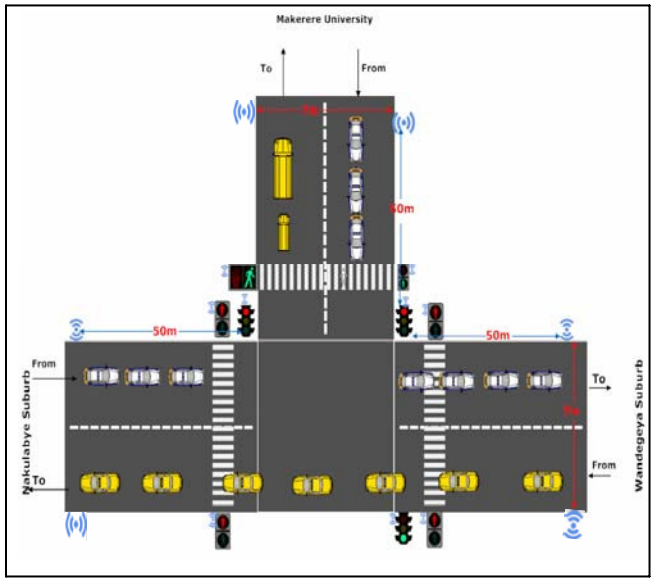



Fig. 3. Proposed Simulator

As depicted in fig.3 above, road traffic simulators should allow a research to graphically overlap a WSN on top of the designed road network. In the above case the blue nodes show the WSN nodes that are placed alongside the lanes forming a T-Junction, the WSN can be used to detect the vehicles on the lane and more efficient models/algorithms for controlling traffic lights basing on the data from the network developed.

### Topology Aware Sensor Nodes

With road traffic simulators, researchers are able to quickly design simple and very complex road networks. If there is need of applying WSNs on such road networks to solve certain ITS problems, the researchers find it hard and time consuming in simulating and adjusting corresponding WSNs. Like from Fig. 3 above, the road traffic simulator should automatically calculate and specify the topology of the deployed sensor nodes according to the pre-designed road network.

### Multiple Virii Sensing

Some ITS require sensor networks that can sensor multiple virii. By virii we mean the kind of data a sensor node is able to sense and send on the network. Some sensor networks support Acoustic, Magnetometer, Light, Temperature, Humidity sensing among others. WSNs simulators should support multiple virii sensing and should easily sense such data from the traffic that is on the designed road network.

### Analysis

Currently road traffic simulators produce analysis of their data and display it separately, so do WSNs simulators even if the two simulators are cooperating to solve the same problem. It will be easier for the researcher to access analysis of WSNs directly from or when within mobility simulators. This avoids switching between two simulators as one carries out the simulation.

## 6. CONCLUSION

This paper has largely highlighted the needed capabilities for both mobility and WSNs simulators. We have also given the overview of current road traffic simulators and WSNs simulators. The aim was to show that the current ITS simulators have tended to be exclusive, domain specific and yet more often than not researchers wish to combine them to solve a common problem. This paper is by no means exhaustive of the needed capabilities for future ITS simulators. Our future work will expound on this and venture more on practically achieving some of the suggested capabilities on some selected simulators.

## 7. REFERENCES


[1] Lars Hoehmann, Anton Kummert, "Mobility Support for Wireless Sensor Networks Simulation for Road Intersection Safety Applications", in *Proceedings of Circuits and Systems, 2009. MWSCAS '09. 52nd IEEE International Midwest Symposium, 2009.*

[2] Green Light District Simulator, http://sourceforge.net/projects/stoplicht/

[3] Simulation of Urban Mobility, http://sourceforge.net/apps/mediawiki/sumo/index.php?title=Main_Page

[5] http://www.tinyos.net/dist-2.0.0/tinyos-2.x/doc/html/tutorial/ lesson1.html

[6] http://www.isi.edu/nsnam/ns/

[7] Malik Tubaishat, Yi Shang, Hongchi Shi, "Adaptive Traffic Light Control with Wireless Sensor Networks", IEEE, 2007

[8] OMNet++, http://www.omnetpp.org/

[9] P. Levis, N. Lee, M. Welsh, D. Culler, "TOSSIM: Accurate and Scalable Simulation of Entire TinyOS Applications," *Proceedings of the First ACM Conference on Embedded Networked Sensor Systems*

[10] jist/swans, http://jist.ece.cornell.edu/.

[11] OPNET, http://www.opnet.com/

[12] E. Brockfeld, P. Wagner, "Testing and Benchmarking of Microscopic Traffic Flow Simulation Models", *In 10th World Conference on Transport Research (WCTR), Istanbul, Turkey, July 2004*

[13] K. Nagel and M. Schreckenberg, "A cellular automaton model for freeway traffic," J. Phys. I France, vol. 2, pp. 2221–2229

[14] http://www.ptvamerica.com/vissim.html

[15]http://www.sias.com/ng/sparamicshome/sparamicshome.htm

[16]http://portal.acm.org/citation.cfm?id=957289.957323

# Multicast routing protocol for ad-hoc networks with route aggregation and transmission power control

Katsuhiro Naito, Michitaka Fujii, Kazuo Mori, and Hideo Kobayashi
Department of Electrical and Electronic Engineering, Mie University,
1577 Kurimamachiya, Tsu, 514-8507, Japan
Email: {naito, kmori, koba}@elec.mie-u.ac.jp

*Abstract*— **In this paper, we propose a multicast routing protocol for ad-hoc networks to reduce packet collisions. Packet collisions are one of the degradation factors in wireless communications. In the ad-hoc multicasting, some forwarder nodes forward data packets from a source node to a lot of destination member nodes. Moreover, forwarding timing of data packets is almost same instance because forwarder nodes with a same hop count receive data packets at same instance from their an upstream node. Therefore, the ad-hoc multicasting tends to suffer from packet collisions due to multiple forwarding of data packets at same instance. In the proposed routing protocol, each forwarder node informs a number of its downstream nodes. Then, neighbor downstream nodes select a forwarder node with maximum number of downstream nodes to aggregate multicast routes. Hence, the proposed protocol can reduce redundant packet forwarding due to a lot of existence of forwarder nodes. Additionally, forwarder nodes control transmission power of data packets according to information from downstream nodes. As the results, the proposed protocol can reduce packet collisions due to hidden terminal problems. From the simulation results, we can find that the proposed protocol can reduce the packet collisions and improver the delivery ratio.**

*Keywords*— **Ad-hoc networks, Routing protocol, Multicast, Collision, Hidden terminal problems, transmission power control**

## I. Introduction

In wireless ad-hoc networks, each node behaves as a router as well as an end host. Hence, end-to-end communication is performed by multi-hop communication. Routing protocols are most important mechanisms for multi-hop communication because packet forwarding is decided by routing protocols. In conventional researches, several routing protocols have been proposed for ad-hoc networks [1].

In ad-hoc networks, various applications have been proposed. Meanwhile, these applications can be categorized into three types: unicast communication, multicast communication, and broadcast communication. In this paper, we focus on the multicast communication because multicast communication is required for new type applications such as a video streaming, an Internet radio, etc. In conventional multicasting researches, several challenges due to changes in network topology and features in wireless communication environment are performed. Since conventional multicast routing protocols for wired networks cannot apply to ad-hoc networks, new routing protocols for ad-hoc networks have been proposed [2], [3].

IEEE 802.11 is a well-known wireless device to achieve ad-hoc networks [4]. IEEE 802.11 has special mechanisms named RTS (Request To Send) / CTS (Clear To Send) for hidden node problems. Meanwhile, RTS/CTS mechanisms are not used for multicast communication because multicast communication employs broadcast modes in IEEE 802.11 systems. Therefore, it is difficult to avoid collisions due to hidden node problems because nodes only perform channel sensing in the broadcast modes [5], [6], [7]. Moreover, some data packets are forwarded hop by hop from a source node. Therefore, some forwarder nodes with a same hop count transmit the data packets at almost same instance. Hence, multicast communication tends to suffer from packet collisions due to hidden node problems. Additionally, almost all multicast routing protocols construct a tree-based topology. As the result, packet collisions at upstream nodes cause many packet loses on their all-downstream nodes in multicast communication[8].

In order to achieve reliable broadcast communication, some Media Access Control protocols for multicasting have been proposed [9], [10]. In these researches, authors extended RTS/CTS mechanism in IEEE 802.11 for broadcast communication. Meanwhile, these extensions require modifications of frame formats or hardware. In addition, few researches of routing protocols have been proposed for solving these issues [11].

In this paper, we propose a multicast routing protocol for ad-hoc networks to reduce packet collisions due to hidden node problems. In the proposed routing protocol, each forwarder node informs a number of its downstream nodes. Then, neighbor downstream nodes select a forwarder node with the maximum number of downstream nodes to aggregate multicast routes. Hence, the proposed protocol can reduce redundant packet forwarding due to a lot of existence of forwarder nodes.

Additionally, forwarder nodes control transmission power of data packets according to information from downstream nodes. As the results, the proposed protocol can reduce packet collisions due to hidden node problems. From the simulation results, we can find that the proposed protocol can reduce the packet collisions and improver the delivery ratio.

## II. ODMRP

In the proposed routing protocol, we employ an On-Demand Multicast Routing Protocol (ODMRP) [12] as a base multicast routing protocol. ODMRP is a well-known mesh based routing protocol for ad-hoc multicasting. ODMRP constructs tree-based routes from source nodes to some destination member nodes. Then, a soft-state approach is taken to maintain multicast members.

When a multicast source node has packets to send, it broadcasts a Join-Query message to an entire network. Join-Query messages are periodically broadcast to inform its multicast service and to refresh membership information and to update routes. When intermediate nodes receive the Join-Query messages, they store a source node address and a sequence number in their message cache to detect any duplicate Join-Query messages. An upstream node address of the Join-Query message is registered into the routing table for a reverse path back to the source node. If Join-Query packets are not duplicate and a Time-To-Live (TTL) value is greater than zero, they will be rebroadcast to neighbor nodes.

When destination nodes, which are multicast member, receive Join-Query messages, they create and transmit a Join-Reply message to their upstream node. When intermediate nodes receive the Join-Reply messages, they check whether their own node address matches the next hop node address within the Join-Reply messages. If they can confirm that the Join-Reply messages are transmitted to own node, they should be forwarding group nodes, which forward data packets from the source node. The Join-Reply messages are propagated by each forwarding group nodes until they reach the multicast source node.

Figure 1 shows the example of route construction in ODMRP. In the example, one source node exists, and ten multicast group member nodes also exist in the network. The locations between node N1 and N4, node N4 and N5, and node N4 and N3 are hidden nodes situation. Therefore, each node cannot sense the other node's signals due to hidden node problems. Solid arrows mean the constructed routes by ODMRP. Dot arrows mean the interference signals from corresponding hidden nodes. Dot circles mean the transmission region of full transmission power.

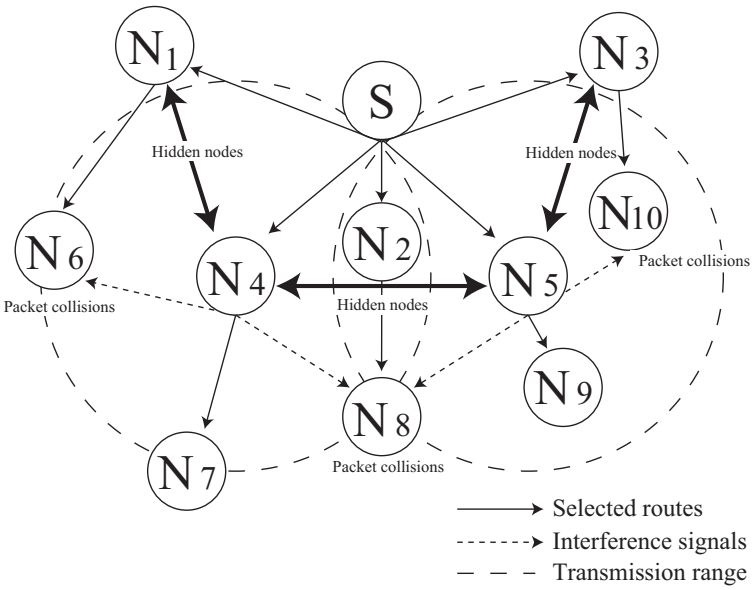



Fig. 1. Example route construction in ODMRP

In ODMRP, nodes select an upstream route for first arriving Join-Query message. As the results, several routes are constructed in the network. In the example, node N6 selects node N1, node N7 selects node N4, node N8 selects node N2, node N9 selects node N5, and node N10 selects node N3 as the upstream node. In data delivery phase, nodes N1, N2, N4, N5 and N3 forward same data packets at almost same instance because the hop count of these nodes is same. Moreover, node N6 can receive signals form nodes N1 and N4, node N8 can receive signals from nodes N2, N4 and N5, and node N10 can receive signals from nodes N5 and N3. As the results, nodes N6, N8 and N10 suffer from packet collisions due to interference from hidden nodes. In multicast delivery, packet collisions at upstream nodes mean packet losses at all corresponding downstream nodes. Therefore, packet collisions at upstream nodes become one of the serious degradation factors in multicast delivery.

## III. Proposed protocol

In order to reduce hidden nodes in ad-hoc networks, we focus on number of FG nodes and transmission area. Hence, the proposed protocol intends to aggregate routes from a source node to reduce the number of FG nodes. Moreover, each node controls transmission power according to desired signal-to-interference and noise power ratio (SINR) to reduce redundant transmission area. As the results, the proposed protocol can reduce the number of hidden nodes.

Meanwhile, the proposed protocol extends the frame formats of Join-Query and Join-Reply messages to carry the

| Type | Reserved | Time To Live | Hop Count |
|---|---|---|---|
| Multicast Group IP Address | | | |
| Sequence Number | | | |
| Source IP Address | | | |
| Last Address | | | |
| Previous Hop IP Address | | | |
| Previous Hop X Coodinate | | | |
| Previous Hop Y Coodinate | | | |
| Previous Hop Moving Speed | | Previous Hop Moving Direction | |
| Minimum Link Expiration Time | | | |
| Number of Downstream Nodes | | | |
| Route Transmission Power | | | |

Fig. 2. Frame format of Join-Query messages

| Type | Count | R | F | Reserved |
|---|---|---|---|---|
| Multicast Group IP Address | | | | |
| Previous Hop IP Address | | | | |
| Sequence Number | | | | |
| Last Address | | | | |
| Required Transmission Power | | | | |
| Hop Count | | | | |
| Sender IP Address[1] | | | | |
| Next Hop IP Address[1] | | | | |
| Route Expiration Time[1] | | | | |
| * * * | | | | |
| Sender IP Address[n] | | | | |
| Next Hop IP Address[n] | | | | |
| Route Expiration Time[n] | | | | |

Fig. 3. Frame format of Join-Reply messages

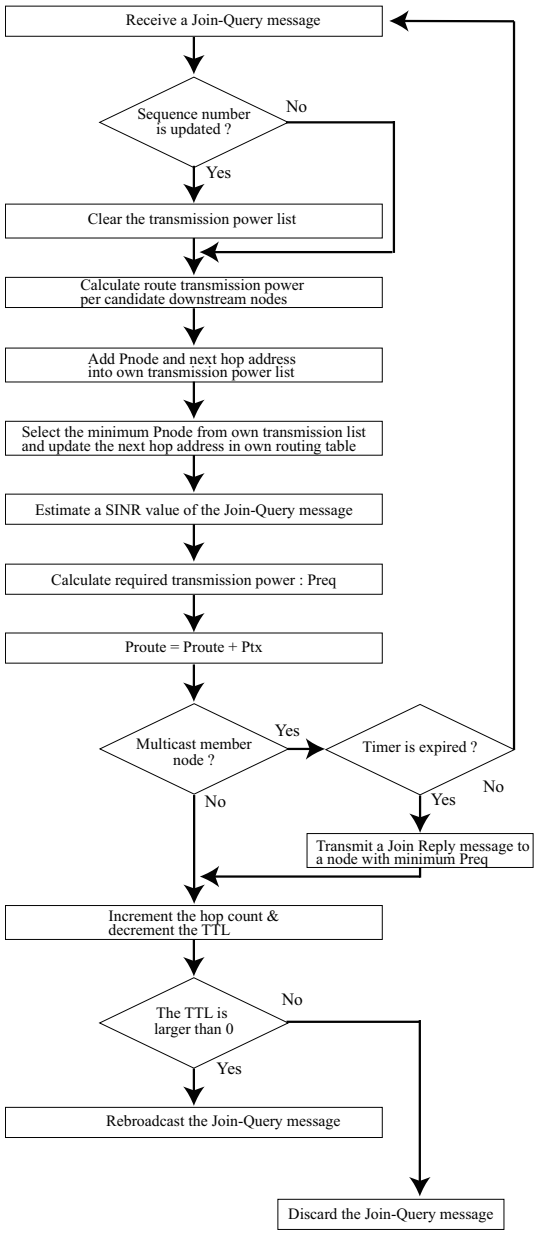



Fig. 4. Flowchart for Join-Query messages

required information. Figure 2 shows the frame format of Join-Query messages. In the extensions, the fields for the number of downstream nodes and the route transmission power are added. Figure 3 shows the frame format of Join-Reply messages. In the frame format, the fields for the required transmission power and the hop count are added.

In order to reduce the number of FG nodes, the proposed protocol employs the number of candidate downstream nodes at FG nodes as the selection criteria of upstream nodes. Figure 5 shows the flowcharts for Join-Reply messages.

When nodes receive Join-Reply messages, they confirm that the own hop count is greater than the hop count of the received Join-Reply message. If the Join reply messages are transmitted from downstream nodes, nodes update the candidate downstream nodes list, which include node address of own downstream nodes. Then, they update the number of candidate downstream nodes, which is informed to neighbor downstream nodes via Join-Query messages.

When nodes confirm that the received Join-Reply messages are transmitted to own node, they should become FG nodes, and rebroadcast the Join-Reply messages to their upstream node with the minimum route transmission power per downstream nodes.

Figure 4 shows the flowchart for Join-Query messages. When intermediate nodes receive Join-Query messages, they

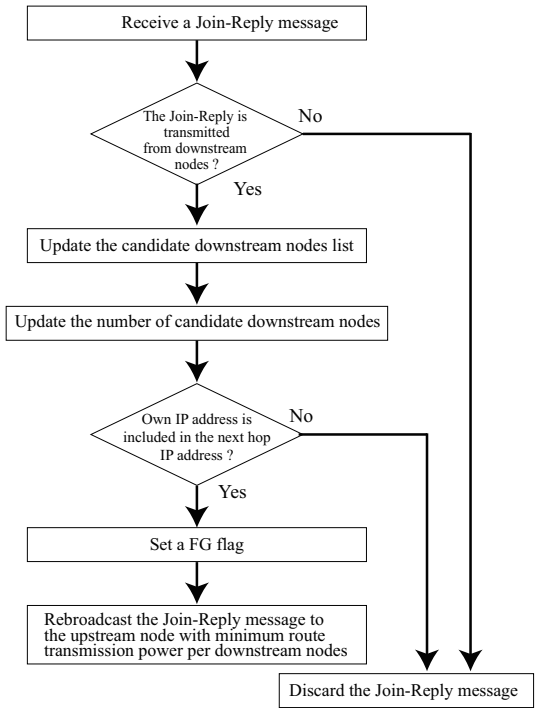



Fig. 5. Flowchart for Join-Reply messages

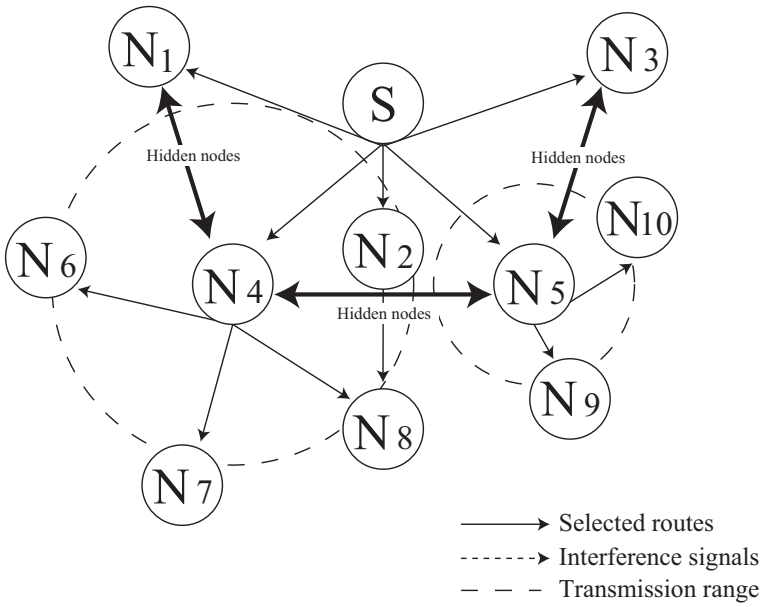



Fig. 6. Example route construction in Proposed protocol

confirm the sequence number. If the sequence number is updated, clear the transmission power list and the node list to refresh the conventional routes. Then, they calculate the route transmission power per candidate downstream nodes $P_{node}$, and add $P_{node}$ and next hop address into own transmission power list. They select the minimum $P_{node}$, which is the most effective forwarding node, and update the next hop address in own routing table. Next, they estimate a SINR value of the received Join-Query messages. Then, they calculate required transmission power according to the estimated SINR value. In order to inform the own route transmission power, they add the own transmission power to the route transmission power in the received Join-Query message. Finally, the received Join-Query messages are rebroadcasted.

When the nodes are multicast member, they collect some Join-Query messages to explore the better candidate upstream nodes for certain period. After the certain period, they transmit a Join-Reply message to their upstream node with the minimum required transmission power $P_{req}$.

## IV. Example operations

Figure 6 shows the example route construction in the proposed protocol. The assumed situation in Fig. 6 is same as Fig. 1. In the proposed protocol, nodes constructs the different route by following procedures.

- The source node S broadcasts Join-Query messages to the whole network. The Join-Query messages include the initial route transmission power $P_{route}(S)$ and the initial number of downstream nodes $N_{down}(S)$. The initial value of $P_{route}(S)$ is the full transmission power of the node S. The initial value of $N_{down}(S)$ sets to 1.
- The nodes N1, N2, N3, N4, and N5 receive the broadcasted Join-Query messages from the source node S. They estimate the SNR of the received Join-Query messages, and calculate the required transmission power $P_{req}$. Additionally, they calculate the route transmission power per candidate downstream nodes $P_{node}$ according to the route transmission power $P_{route}(S)$ and the number of downstream nodes $N_{down}$. Then, they register $P_{node}$, $P_{req}$, and corresponding upstream node into the transmission power list.
- The nodes N1, N2, N3, N4, and N5 select the minimum $P_{node}$ from own transmission list and update the next hop address in own routing table. Then, they reply the Join-Reply messages to the selected upstream node when they are member of multicasting.
- The nodes N1, N2, N3, N4, and N5 calculate the new route transmission power $P_{route}$ by adding $P_{route}(S)$ and the own transmission power $P_{tx}$. For example, $P_{route}(N1) = P_{route}(S) + P_{tx}(N1)$. Then, they broadcast the own Join-Query messages including the new route transmission power $P_{route}$ and the own number of downstream nodes $N_{down}$.
- The nodes N6, N7, N8, N9 and N10 receive the Join-Query messages from the nodes N1, N2, N3, N4 and

N5. They proceed in the same operations such as the operations of the nodes N1, N2, N3, N4 and N5. Finally, they reply the Join-Reply messages to the upstream node.

- The nodes N1, N2, N3, N4 and N5 can receive the Join-Reply messages from the nodes N6, N7, N8, N9 and N10. Therefore, they can count the candidate number of downstream nodes, and update $N_{down}$. Additionally, nodes, that receive the Join-Reply messages including own address, can update the own transmission power $P_{tx}$ according to the $P_{req}$ of the downstream nodes. If the number of downstream nodes is not one, they select the maximum $P_{req}$ for the own transmission power $P_{tx}$.
- The nodes N1, N2, N3, N4 and N5 can update the route transmission power $P_{route}(S)$ when the source node rebroadcast the next Join-Query messages. As the results, $P_{node}(N4)$ and $P_{node}(N5)$ are less than $P_{node}(N1)$, $P_{node}(N2)$ and $P_{node}(N3)$. Therefore, nodes N6, N7 and N8 can select the node N4 as their upstream node, and nodes N9 and N10 can select the node N5 as their upstream node. Finally, multicast routes can be aggregated through the nodes N4 and N5. Then, packet collisions at the nodes N6 and N10 can be avoided according to the route aggregation.
- In the data delivery phase, the nodes N4 and N5 transmit data packets with the minimum transmission power $P_{tx}(N4)$, and $P_{tx}(N5)$. As the results, the transmission regions of the nodes N4 and N5 are reduced like as Fig. 6. Then, packet collisions at the nodes N8 can be avoided according to the transmission power control.

## V. Numerical results

In this section, we compare the performance for the proposed protocol with that for the conventional ODMRP protocol. The simulations are performed by the network simulator QualNet[13]. In the simulations, we assume the IEEE 802.11g as the wireless communication device, and the transmission rate is fixed at 54 [Mbps]. 100 nodes are placed randomly in 1000 $\times$ 1000 [m] area. The source and the member nodes are selected randomly. The application is CBR (Constant Bit Rate) and data packets with the length of 1 [KB] are transferred for 300 [s]. We consider the additive white gaussian noise (AWGN) environment and the free space propagation model. A target SINR in the proposed protocol is set to 20 [dB]. The simulation results are an average of 100 simulation trials.

Figure 7 shows the number of collisions per member nodes. From the results, we can find that our proposed protocol can reduce the packet collisions. This is because, the proposed protocol can reduce FG nodes by aggregating multicast delivery routes and control transmission power to avoid hidden node problems.

TABLE I

Simulation parameters

| | |
|---|---|
| Simulator | Qualnet 4.5 |
| Simulation period | 300 [s] |
| Simulation area | 1000 $\times$ 1000 [m] |
| Number of nodes | 100 |
| Node position | Random |
| Mobility | None |
| Data packet length | 1048 [byte] |
| Transmission interval | 65 [ms] |
| Wireless device | IEEE 802.11g |
| Transmission rate | 54 [Mbps] |
| Wireless environment | Free space |
| Application | CBR 128 [kbps] |
| Routing protocol | ODMRP, Proposed |
| Simulation trials | 100 |
| Buffer period of Join-Query messages | 10[ms] |
| Target SINR | 20[dB] |

Figure 8 shows the packet delivery ratio at member nodes. From the results, the proposed protocol can improve the packet delivery ratio because our protocol can reduce hidden nodes and improve performance degradation due to hidden node problems.

Figure 9 shows the number of route change per data packets. From the results, we can find that the proposed protocol can also reduce the number of route change. This is caused by the aggregation mechanisms of the proposed protocol. On the contrary, ODMRP select some neighbor nodes randomly. Therefore, the number of route change in ODMRP increases.

Figure 10 shows the average number of FG nodes in the network. From the results, we can find that the number of FG nodes is reduced in the proposed protocol. The reason is that the proposed protocol can aggregate multicast route according to the route transmission power per candidate downstream nodes.

## VI. Conclusions

This paper proposed the new multicast routing protocol for ad-hoc networks. The proposed protocol is based on the ODMRP, which is the most well-known multicast protocol. The features of the proposed protocol are to improve hidden nodes problems by aggregating multicast routes and controlling transmission power. From the simulation results, we confirmed that the hidden node problem is the one of the degradation factors for ad-hoc multicasting, and our protocol can improve the packet delivery ratio and route stability.

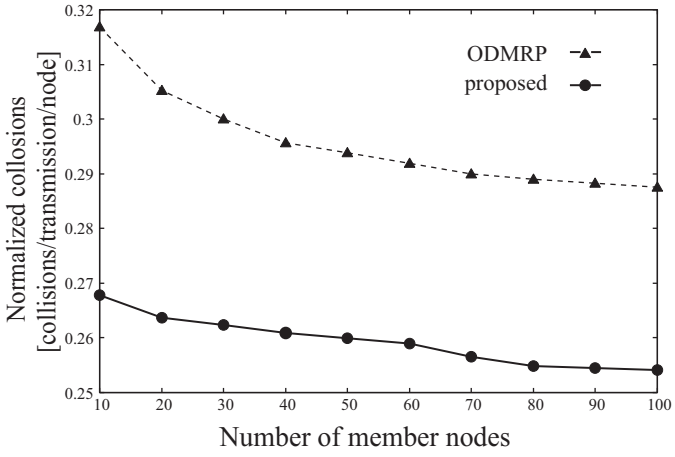



Fig. 7. Normalized collisions

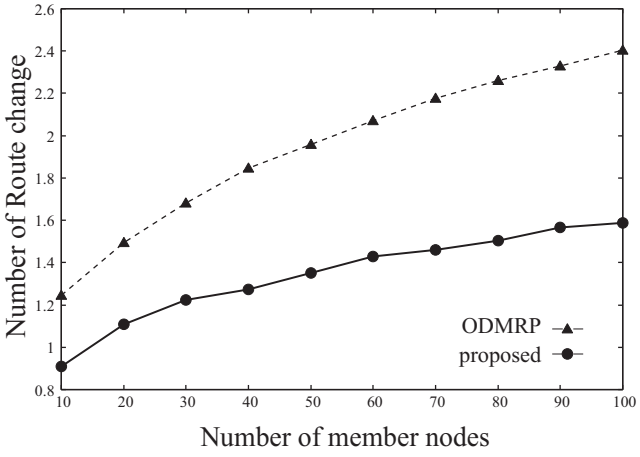



Fig. 9. Normalized number of route changes

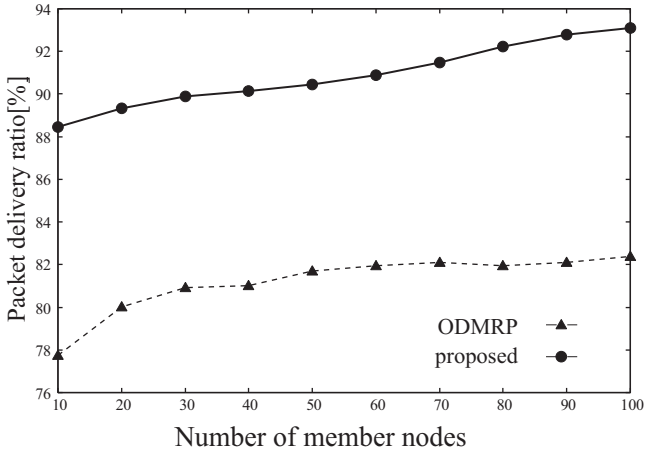



Fig. 8. Packet delivery ratio

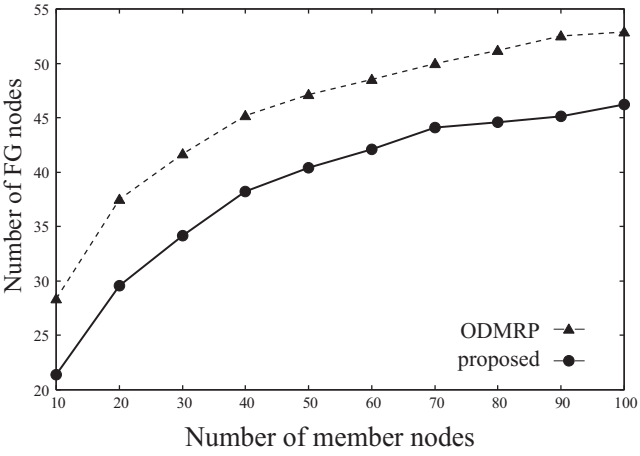



Fig. 10. Average number of FG nodes

## ACKNOWLEDGMENT

This work was supported by the Telecommunications Advancement Foundation and MEXT KAKENHI(23700075).

## REFERENCES

[1] H. Y. Hsieh and R. Sivakumar, "Performance comparison of celluar and multi-hop wireless networks: A quantitative study," Proc. ACM SIGMETRICS, pp. 113–122, Boston, MA, USA, Jun. 2001.

[2] C. d. M. Cordeiro, H. Gossain, and D. P. Agrawal, "Multicast over Wireless Mobile ad hoc networks: Present and Future Direction," IEEE Network January/February 2003.

[3] K. Viswanath, K. Obraczka, and G. Tsudik, "Exploring Mesh and Tree-based Multicast. Routing Protocols for MANETs," IEEE Transactions on Mobile Computing, vol. 5, no. 1, pp. 28–42, Jan. 2006.

[4] K. Tang, and M. Gerla, "Random Access MAC for Efficient Broadcast Support in ad hoc networks," IEEE Wireless Communications and Networking Conference (WCNC 2000), pp. 454–459, Sept. 2000.

[5] S. Gupta, V. Shankar, and S. Lalwani, "Reliable Multicast MAC Protocol for Wireless LANs," IEEE ICC 03, vol. 1, pp. 93–97, 2007.

[6] W. Si, and C. Li, "RMAC: A Reliable Multicast MAC Protocol for Wireless adhoc networks," IEEE International Conference on Parallel Processing (ICPP 04), vol. 1, pp. 494–501, Aug. 2004.

[7] C.Y. Chiu, E.H.-K. Wu, Y. Wang, and G.-H. Chen, "A Reliable and Efficient MAC Layer Broadcast Protocol for Mobile ad hoc networks," IEEE Transactions on Vehicular Technology, vol.56, no.4, pp.2296–2305, Jul. 2007.

[8] X. Wang, L. Wang, Y. Wang, and D. Gu, "Reliable Multicast Mechanism in WLAN with Extended Implicit MAC Acknowledgment," IEEE Vehicular Technology Conference 2008, pp. 2695–2699, 2008.

[9] K. Lee and D. Cho, "A multiple access collision avoidance protocol for multicast service in mobile ad hoc networks," IEEE Communications Letters, vol. 7, no. 10, pp. 508–510, Oct. 2003.

[10] K. S. Lau and C. W. Pao, "Collision Avoidance and Recovery for Multicast Communications in ad hoc networks," IEICE transactions on communications, vol. E88-B, no.7 pp. 2937–2943, Jul. 2005.

[11] M. Pandey and D. Zappala, "A Scenario-Based Evaluation of Mobile Ad Hoc Multicast Routing Protocols," Jul. 2004.

[12] Sung-Ju Lee, William Su, Mario Gerla, "On-Demand Multicast Routing Protocol (ODMRP) for ad hoc networks," Internet-Draft, draft-ietf-manet-odmrp-02.txt, Jan. 2000.

[13] Network Simulator Qualnet is http://www.scalable-networks.com/

# Proposal of traffic aware routing based on neighborhood communication for ad-hoc networks

Katsuhiro Naito, Kazuto Toyozumi, Kazuo Mori, and Hideo Kobayashi
Department of Electrical and Electronic Engineering, Mie University,
1577 Kurimamachiya, Tsu, 514-8507, Japan
Email: {naito, kmori, koba}@elec.mie-u.ac.jp

*Abstract*— **In this paper, we propose a traffic aware routing protocol for ad-hoc networks. In the proposed protocol, nodes estimate the utilizable wireless resource for constructing routes. Moreover, they consider the neighbor traffic according to the route requested connection. In order to achieve the above objects, nodes always sense a wireless channel to confirm the channel condition. Then, they remove the traffic of route requested connection from measured data to estimate the utilizable wireless resource without own connection traffic. To construct routes based on the measured traffics, nodes select forwarding delay period according to the utilizable wireless resource. From simulation results, we can find that the proposed protocol can improve the packet delivery ratio according to dispersed traffic.**

*Keywords*— **Ad-hoc networks, Traffic aware, Interference reduction, Routing protocol**

## I. Introduction

In wireless ad-hoc networks, same frequency band is shared by all wireless nodes. Therefore, sharing mechanisms of wireless resource are important factors to achieve ad-hoc networks. Carrier sense multiple access (CSMA) is the most well-known mechanisms for media access control. If nodes employ the CSMA, they sense a wireless channel first when they try to transmit a data packet. Then, they transmit the data packet when the wireless channel is not used.

In the ad-hoc networks, some nodes may try to communicate simultaneously. Meanwhile, nodes cannot transmit a data packet due to CSMA mechanisms when neighbor nodes are transmitting data packets. Therefore, traffic distribution is an important factor to realize simultaneous communication in ad-hoc networks.

The conventional routing protocols for ad-hoc networks are classified into three categories: minimum hop count protocols, consumed power aware protocols, and traffic aware protocols [1]. Ad Hoc On-demand Distance Vector Routing (AODV) [2] and Dynamic Source Routing (DSR) [3], [4] are well-known protocols in the minimum hop count category. The minimum hop count routing protocols can construct minimum distance routes. Meanwhile, the performance degrades due to traffic concentration. Consumed power aware routing protocols can construct routes with minimum consumed power for communication [5], [6], [7]. Therefore, it is useful to extend lifetime of wireless nodes. Traffic aware routing protocols can construct routes with low traffic [8]. Hence traffic distribution can be achieved, and throughput is also improved.

In this paper, we focus on the traffic aware routing protocols to achieve reliable simultaneous communication in ad-hoc networks. In the conventional traffic aware routings, some factors such as number of transfered data packets [9], number of links to neighbor nodes [10], [11], busy ratio of physical wireless channel [12], [13], etc. are considered as an indicator of traffic.

Moreover, collecting schemes of traffic information are classified into two categories. First one is to collect all traffic information in whole networks [9], [10], [11], [12], [13]. In this category, nodes can find an optimum route to detour heavy traffic area. On the contrary, an overhead for collection of traffic information becomes especially large. Second one is to collect partial traffic information in whole networks [14], [15], [16], [17], [18]. In this category, nodes can find a sub optimum route. But, an overhead will be smaller than that of first category.

In CSMA mechanisms, nodes transmit data packets when they confirm that wireless channel is not used by neighbor nodes. Moreover, wireless signals spread in a circular pattern at the center of a transmitting node. Therefore, to detour heavy traffic area, nodes should confirm the wireless channel status.

In multi-hop communication, neighbor traffic of own connection affect to own communication, because some data packets are forwarded hop by hop over the route [19], [20]. Additionally, routes may be reconstructed due to link losses, movement of nodes, etc. in ad-hoc networks. Therefore, routing protocols should consider the traffic of own connection not to detour own traffic.

In this paper, we propose a new traffic aware routing protocol for ad-hoc networks. The proposed protocol employs the channel busy ratio, which is a utilizable ratio of wireless

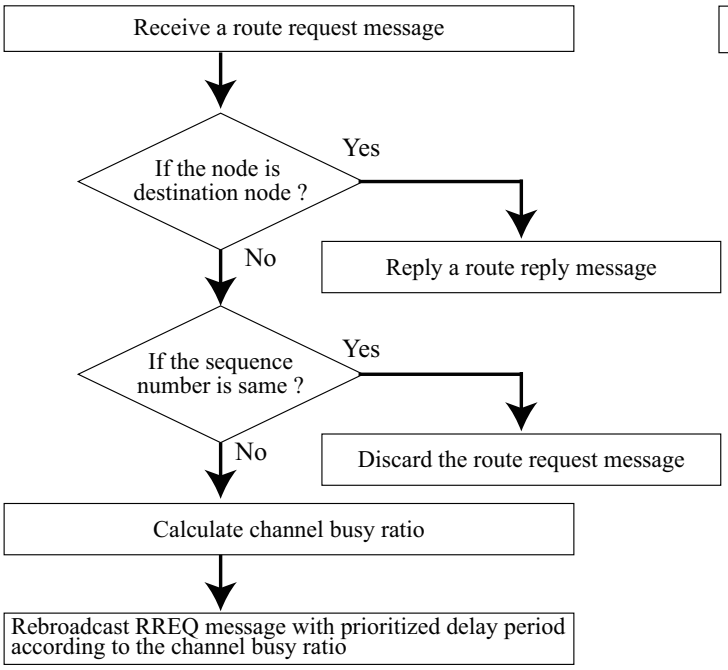



Fig. 1. Flowchart for route request messages.

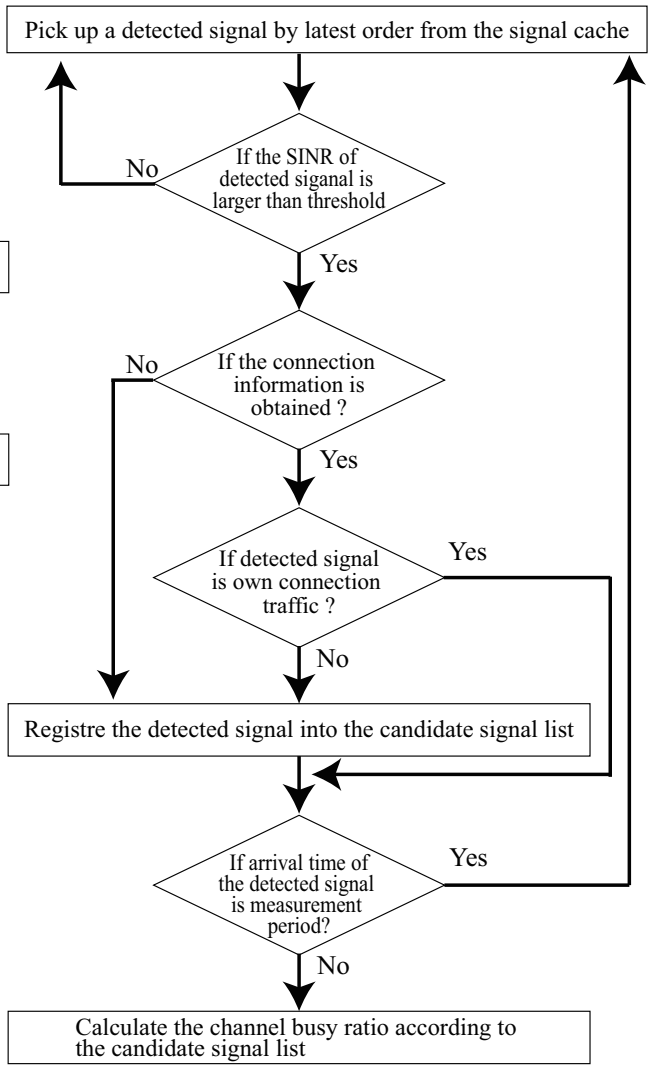



Fig. 2. Flowchart for calculation of channel busy ratio.

channel, as an indicator of traffic. Then, nodes consider the own connection traffic to calculate the channel busy ratio. Meanwhile, the proposed protocol employs AODV as the base routing protocol. Therefore, nodes achieve prioritized route construction by changing forwarding delay period according to the channel busy ratio. From numerical results, we can find that the throughput performance can be improved when simultaneous communication is performed.

## II. Proposed traffic aware routing

### A. Prioritized route construction

The proposed routing protocol is designed based on the AODV protocol. AODV is one of the reactive routing protocols for ad-hoc networks. Nodes transmit a route request (RREQ) message to whole networks when the nodes try to transmit data packets. Neighbor nodes rebroadcast the received RREQ message to achieve flooding. A destination node replies a route reply (RREP) message to the source node when it receives the RREQ message to own node.

Figure 1 shows the flowchart for route request messages. When nodes receive a route request message, they check the destination address in the route request message. If the destination address matches own address, nodes reply a route reply message to the host of the source address in the route request message. If not, nodes check the sequence number in the route request message. If the sequence number is updated, nodes start to calculate the channel busy ratio. In the route construction of AODV, destination nodes select a route, which conveys the first arrival route request message. Therefore, the proposed protocol controls the forwarding delay period of the route request messages. Hence, nodes set the forwarding delay period of the route request message according to the calculated channel busy ratio. In the procedure of prioritization, long delay period is set for high channel busy ratio not to be selected as a route.

### B. Calculation algorithms for channel busy ratio

The feature of the proposed scheme is to consider the own connection traffic to calculate channel busy ratio(CBR). Figure 2 shows the flowchart for calculation of channel busy ratio. In the proposed protocol, nodes are always sensing a wireless channel to collect traffic information. The collected traffic information is registered into the signal cache. When nodes start to calculate a channel busy ratio, nodes pick up detected signals by latest order from the signal cache for

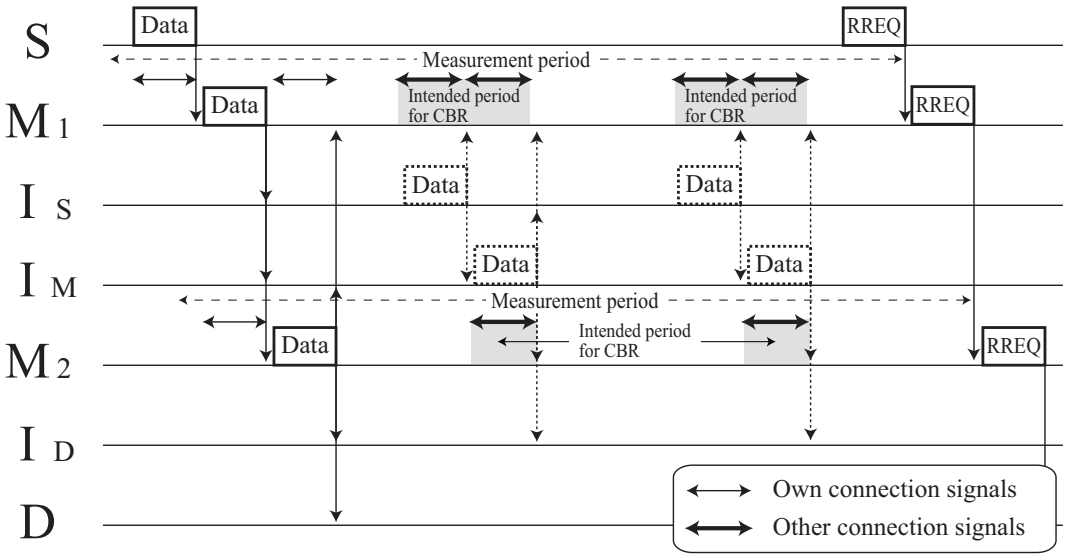



Fig. 4. Example calculation of channel busy ratio in packet flow.

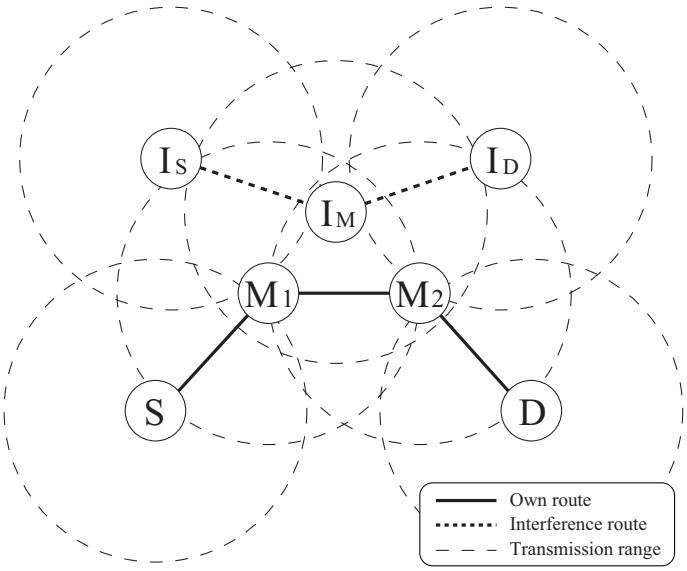



Fig. 3. Example network for calculation of channel busy ratio.

certain period. If the SINR of the detected signal is larger than the sensing threshold in IEEE 802.11 systems, nodes try to check the connection information such as a source address and a destination address in the packet. When the connection information is obtained, nodes checks the detected signal is a part of connection, which is requested in the received route request message, through the own node. If the detected signal is other traffic, nodes register the detected signal into the candidate signal list. Above-described processes are repeated for certain measurement period. Finally, nodes calculate the channel busy ratio from the candidate signal list.

### C. Example calculation of channel busy ratio

Figure 3 shows the example network for channel busy ratio, and Fig. 4 shows the example calculation of channel busy ratio. In the example network, node $I_S$ communicates with node $I_D$ through node $I_M$. Node S has communicated with node D through node $M_1$ and $M_2$, and it tries to find a new route to node D due to a route failure. Node $M_1$ exists in the transmission ranges of node $I_S$ and node $I_M$, and node $M_2$ exists in the transmission ranges of node $I_M$ and node $I_D$.

In Fig. 4, node S transmits the data packet to node D through node $M_1$ and node $M_2$. Therefore, node $M_1$ and node $M_2$ receive the data packets. Then, node $I_S$ transmits the two data packets to node $I_D$. Since node $M_1$ and node $M_2$ exist in the transmission regions, they receive the data packets of the communication between node $I_S$ and node $I_D$.

When node $M_1$ receives the RREQ message from node S, it starts to calculate the channel busy ratio for the certain measurement period. In the example, first data packet is ignored for the channel busy ratio because it is the data packet of own connection from node S. On the contrary, successive four data packets are targeted for the channel busy ratio because they are the data packets of other connection. By the same token, node $M_2$ can calculate the channel busy ration for the the certain measurement period.

## III. Example operations

Figure 5 shows the example route construction procedures, and Fig. 6 shows the packet flow of routing control messages. In the example, constant bit rate with 128 [Kbps] is performed between node $I_S$ and $I_D$ as the interference connection. Additionally, bandwidth is assumed to 2 [Mbps]. When node

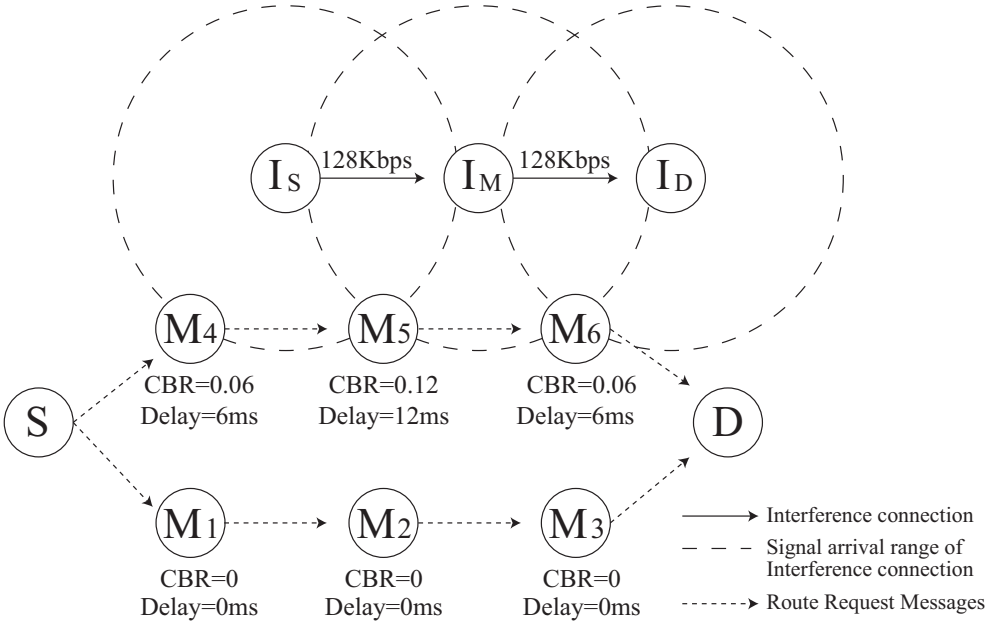



Fig. 5. Example route construction procedures.

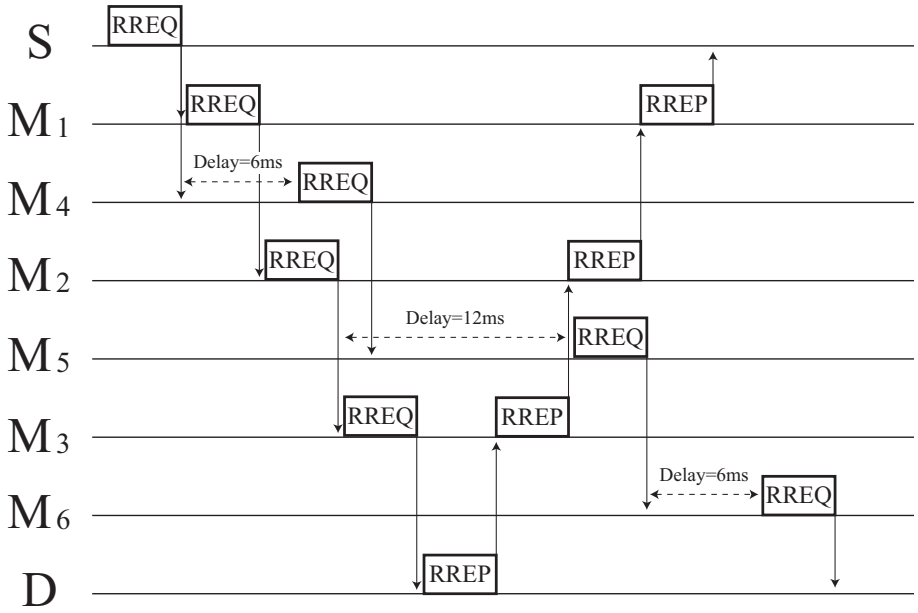



Fig. 6. Example flow of control messages.

S tries to construct a route to node D, the following procedures are performed.

- Node S broadcasts the RREQ message to whole network in order to find routes to node D.
- Node $M_1$ and node $M_4$ receive the broadcasted RREQ message from node S. They calculate the channel busy ratio, and pick up detected signals by latest order from the signal cache for certain period. Node $M_1$ does not receive any signals from neighbor nodes. Therefore, the channel busy ratio (CBR) of node $M_1$ is 0. On the contrary, node $M_4$ receives signals from node $I_S$. Since node $I_S$ transmits data with 128 [Kbps] and the bandwidth is 2 [Mbps], the CBR of node $M_4$ is 0.06. Then, they set the forwarding delay according to the CBR. Therefore, the forwarding delay of node $M_1$ is set to 0 [ms], and the forwarding delay of node $M_4$ is set to 6 [ms]. As the results, node $M_1$ forwards the RREQ message without the forwarding delay. Meanwhile, node $M_4$ forwards the

TABLE I

FORWARDING DELAY PERIOD.

| Channel busy ratio(CBR) | Delay [ms] |
|---|---|
| 0 - 0.2 | 0 - 20 |
| 0.21 - | 600 |

TABLE II

SIMULATION PARAMETERS.

| | |
|---|---|
| Simulator | QualNet |
| Simulation time | 500 [s] |
| Simulation trials | 50 times |
| Number of nodes | 200 |
| Simulation area | 11000 × 11000 [m] |
| Node placement | Uniform |
| Node mobility | None |
| Communication system | IEEE 802.11g |
| Transmission rates | 6[Mbps] |
| Propagation pathloss model | Free space |
| Wireless environment | AWGN |
| Routing protocol | AODV, Proposed protocol |
| Application | CBR 64 [Kbps] |
| Data packet size | 1 [KB] |
| Call arrival rate, | 1/20 - 1 |
| Average call-holding time, | exponential with 20 [s] |
| Measurement period, | 3 [s] |

RREQ message with the 6 [ms] forwarding delay.

- Node $M_2$ receives the RREQ message from node $M_1$, and node $M_5$ receives the RREQ message from node $M_4$. Node $M_2$ does not receive any signals from neighbor nodes. Therefore, the channel busy ratio (CBR) of node $M_2$ is 0. On the contrary, node $M_5$ receives signals from node $I_S$ and node $I_M$. Therefore, the CBR of node $M_5$ is 0.12. As the results, the forwarding delay of node $M_2$ is set to 0 [ms], and the forwarding delay of node $M_5$ is set to 12 [ms].
- Node $M_3$ receives the RREQ message from node $M_2$, and node $M_6$ receives the RREQ message from node $M_5$. By the same token, they set the forwarding delay according to the CBR. Therefore, the forwarding delay of node $M_3$ is set to 0 [ms], and the forwarding delay of node $M_5$ is set to 6 [ms].
- Node D receives the RREQ message from node $M_3$ earlier than the RREQ message from node $M_6$. Therefore, it replies the RREP message to node S through node $M_3$, $M_2$, and $M_1$. Finally, the route without neighbor traffic can be constructed without exchange of traffic information between nodes.

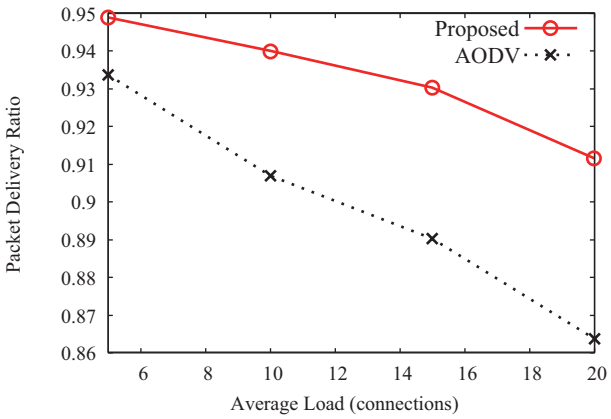



Fig. 7. Packet delivery ratio.

## IV. NUMERICAL RESULTS

In this section, we compare the performance for the proposed protocol with that for the conventional AODV protocol. The simulations are performed by the network simulator QualNet[21]. In the simulations, we assume IEEE 802.11g as the wireless communication device, and the transmission rate is fixed at 6 [Mbps]. 200 nodes are placed uniformly in 11000 × 11000 [m] area. The source and the destination node are selected randomly. The application is constant bit rate with 64 [Kbps] and data packets with the length of 1 [KB]. We consider the additive white gaussian noise (AWGN) environment and the free space propagation model. The simulation results are an average of 50 simulation trials. Detail simulation parameters are shown in Table II.

Figure 7 shows the delivery ratio of data packets. From the results, we can find that the proposed protocol can keep the higher delivery ratio even if the number of connection increases. The reason is that the proposed protocol can select routes in unoccupied area by prioritized route construction according to the channel busy ratio. On the contrary, the performance of AODV degrades with increasing in the number of connections. This is because, nodes construct a minimum hop route in AODV, and some constructed route may be overlapped. Therefore, neighbor connections affect each other.

Figure 8 shows the normalized link failures per data packets. The results show that the proposed protocol can reduce the link failures. When nodes employ the IEEE 802.11 systems, route failures are usually detected by transmission failure in data-link layer. Therefore, interference between neighbor traffics causes the increasing of link failures.

Figure 9 shows the normalized number of route request messages per data packets. From the results, we can find that the proposed protocol can reduce the number of route request messages. This is because, our protocol can reduce the number

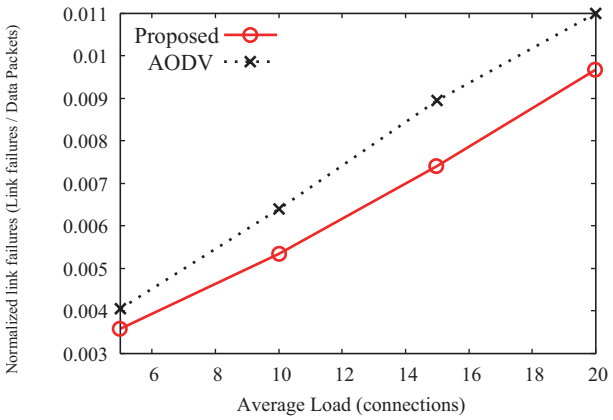



Fig. 8. Normalized link failures per data packets.

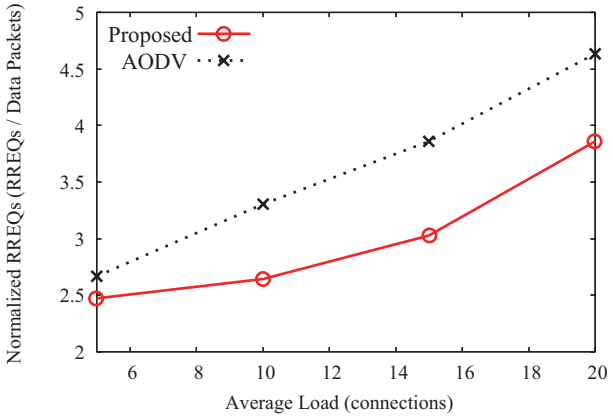



Fig. 9. Normalized number of route request messages per data packets.

of link failures by detouring heavy traffic area. Therefore, route reconstruction is also reduced in the proposed protocol. On the contrary, nodes try to reconstruct routes in AODV, and the number of route request messages also increases.

## V. Conclusions

The traffic aware routing is one of the performance improvement schemes for ad-hoc networks. Especially, communication in ad-hoc networks performs simultaneously. Moreover, same connection traffic affects each other in ad-hoc networks. In this paper, we proposed the traffic aware routing, which focus on the unused wireless channel resource and traffic of own connection. From the simulation results, we confirmed that the interference between connections is the one of the degradation factors for ad-hoc networks, and our protocol can improve the throughput performance with a few control messages.

## Acknowledgment

This work was supported by the Telecommunications Advancement Foundation and MEXT KAKENHI(23700075).

## References

[1] J. Broch, D. A. Maltz, D. B. Johnson, Y. C. Hu and J. Jetcheva, A performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols, Proc. of the ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom), pp. 85–97, Oct. 1998.

[2] C. E. Perkins, E. M. Royer and S. R. Das, Ad hoc On-Demand Distance Vector (AODV) Routing, RFC-3561, IETF Network Working Group, 2003.

[3] D. Johnson, Y. C. Hu and D. Maltz, The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4, RFC-4728, IFTF Network Working Group, 2007.

[4] S. R. Das, C. E. Perkins, and E. M. Royer, Performance comparison of two on-demand routing protocols for ad hoc networks, IEEE Conference on Computer Communications, Vol. 1, pp. 3–12, 2000.

[5] X Jing and M. J. Lee, Energy-Aware Algorithms for AODV in Ad Hoc Networks, Proc. Mobile Computing and Ubiquitous Networking, 2004.

[6] E. O. Ochola, D. Chatelain and A. Kurien, Dynamic Power Control On-demand Routing Protocol, Southern African Telecommunications and Networks Conference, 2007.

[7] A. Rani and M. Dave, Load Balanced Routing Mechanisms for Mobile Ad Hoc Networks, Communication, Network and System Sciences, pp. 627–635, Oct. 2009.

[8] A. Rani and M. Dave, Performance Evaluation of Modified AODV for Load Balancing, Journal of Computer Science, Science Publications, pp. 863–868, Nov. 2007.

[9] S. Lee and M. Gerla, Dynamic load-aware routing in ad hoc networks, IEEE Int' l Conference on Communications, Vol. 10, pp. 3206–3210, 2001.

[10] V. Saigal, A. K. Nayak, S. K. Pradhan, and R. Mall, Load balanced routing in mobile ad hoc networks, Elsevier Computer Communications, Vol. 27, pp. 295–305, Feb. 2004.

[11] H. Hassanein, and A. Zhou, Routing with Load Balancing in Wireless Ad Hoc Networks, Proc. ACM MSWiM, Rome, Italy, pp. 89–96, 2001.

[12] J. Hakoda, H. Uehara and M. Yokoyama, Performance Evaluation of Mobile Ad Hoc Routing Protocols Based on Link Expiration Time and Load of Node, IEICE Trans. Commun. , Vol. J85B, No. 12, pp. 2108–2118, 2002.

[13] S. Takahashi, J. Hakoda, H. Uehara and M. Yokoyama, A Load Balanced Routing Scheme for Mobile Ad Hoc Networks, ISITA, Parma, Italy, Oct. 2004.

[14] J. Song, V. Wong, and V. Leung, Load-aware on-demand routing (LAOR) protocol for mobile ad hoc networks, Proc. IEEE Vehicular Technology Conference (VTC- Spring), Vol. 3, pp. 1753–1757, Apr. 2003.

[15] Y . J. Lee, and G. F. Riley, A Workload-Based Adaptive Load-Balancing Technique for Mobile Ad Hoc Networks, IEEE Communication Society, WCNC, Vol. 4, pp. 2002–2007, Mar. 2005.

[16] S. Jung, Nisar Hundewale and Alex Zelikovsky, Energy Efficiency of Load Balancing in MANET Routing Protocols, SNPD/SAWN, pp. 476–483, May 2005.

[17] Y. S. Kim, Y. J. Cho, K. Kang, An Enhanced Load-balancing Approach Using Detour in Ad-hoc Networks, Wireless Communications and Networking Conference, pp. 1–5, Apr. 2009.

[18] J. H. Song, V. W. S. Wong and V. C. M. Leung, Efficient on-demand routing for mobile ad hoc wireless access networks, IEEE journal on selected Areas in Communications, pp. 1374–1383, Sep. 2004.

[19] R. Hekmat and P. Mieghem, "Interference in wireless multi-hop ad-hoc networks and its effect on network capacity, " ACM Wireless Networks, vol 10 , no. 4, pp. 389–399, Jul. 2004.

[20] L. Ma, Q. Zhang, and X. Cheng, "A power controlled interference aware routing protocol for dense multi-hop wireless networks, " AMC Wireless Networks, vol. 14, no. 2, pp. 247–257, Mar. 2008.

[21] Network Simulator QualNet, http://www.scalable-networks.com

# Evaluation of reliable multicast delivery systems with base station diversity and forward error correction

Katsuhiro Naito, Kazuo Mori, and Hideo Kobayashi
Department of Electrical and Electronic Engineering, Mie University,
1577 Kurimamachiya, Tsu, 514-8507, Japan
Email: {naito, kmori, koba}@elec.mie-u.ac.jp

***Abstract*— This paper proposes a reliable multicast delivery system using base station diversity and error-correcting codes for cellular systems. Conventional cellular systems use single wireless communication to achieve multicast delivery. Additionally, they retransmit lost frames to realize reliability of communication. In cellular systems, received signal intensity declines in cell edge areas because distance between a base station and a wireless terminal becomes long. Hence, wireless terminals in cell edge area suffer from many transmission errors due to low received signal intensity. The proposed system utilizes multiple wireless link communication and error-correcting codes to improve transmission performance. The utilizing multiple wireless links improve packet error ratio, and error-correcting codes contribute to reduce a number of retransmission frames and delivery delay. In the proposed system, base stations transmit encoded frame data by error-correcting codes. Each wireless terminal communicates with some neighbor base stations, and receives frame data by combining some frames from different neighbor base stations. Then, it decodes the received frame data by error-correcting codes. As the results, the proposed system can utilize the diversity effect by using multiple base stations and frame data recovery by error-correcting codes. The numerical results show that the proposed system can achieve reliable multicast data delivery with a short transmission period and can reduce consumed wireless resource due to frame retransmission.**

***Keywords*— Cellular systems, Multicasting, Base station diversity, Error-correcting codes**

## I. Introduction

Multicast communication, that delivers same data to many terminals, has drawn attention with widespread of the Internet. Reliable multicast is the one of the recent research topics for multicast communication to achieve reliable data delivery. Some researchers proposed retransmission schemes, in which a source terminal and forwarding terminals retransmit lost data packets[1]. Additionally, some researchers employ error-correcting codes to recover transmission failures at receiver terminals[2].

Recently, reliable multicast communication in wireless networks has been studied according to development of wireless communication technologies[3], [4], [5], [6], [7]. A base station can deliver same data packets to many wireless terminals with single transmission in wireless networks. On the contrary, wireless terminals suffer from packet losses due to transmission errors over wireless links. Packet losses due to transmission errors occur independently at wireless terminals because each wireless link between a base station and a wireless terminal has different characteristics. As the results, a base station should retransmit the same data packets repeatedly due to several requests from wireless terminals. Therefore, wireless resource is also consumed and communication performance is also degraded due to redundant retransmission of the data packets.

A base station usually controls transmission power to deliver data packets to wireless terminals in general cellular networks. However, the base station should transmit data packets with maximum transmission power in multicast communication because many wireless terminals exist in whole cell area. As the results, wireless terminals in cell-edge areas suffer from weak signal power compared with wireless terminals in cell-center areas. Therefore, wireless terminals in cell-edge areas request retransmissions of data packets several times.

Micro-cellular systems which operate small scale communication areas have been considered for high-speed and low-power operation[8]. In micro-cellular systems, cell areas are small in size and many base stations are provided. Therefore, wireless terminals should be able to receive some signals from multiple base stations. Additionally, wireless terminals can combine some signals from different base stations, and can improve transmission performance[9], [10], [11]. Especially, wireless terminals in cell-edge areas tend to improve transmission performance because they receive some signals from different base stations with almost equal received power. Therefore, authors have been studied about base station diversity systems to improve multicast communication performance[12]. From the previous studies, we can

confirm that the base station diversity is effective mechanisms to achieve reliable multicasting in cellular systems. On the contrary, the number of retransmission frames is still large number because many wireless terminals request different frame retransmission.

In this paper, we propose a new base station diversity system that employs error-correcting codes for multicast communication. There have already been studies of error-correcting codes for multicast communication in wired networks[1]. In wired networks, same IP datagram is generally lost because IP datagrams are lost due to congestion. Therefore, they employ error-correcting codes for many IP datagrams to recover lost IP datagrams. Then, receivers should wait for receptions of all IP datagrams which are encoded by error-correcting codes. As the results, the increase of waiting period increases delivery delay. Additionally, performance with base station diversity and error-correcting codes has not been considered in conventional works.

In wireless networks, different frames are lost according to channel condition between a base station and a wireless terminal. Additionally, some wireless terminals request different frame retransmission. Therefore, base stations should retransmit almost all frames to wireless terminals to recover lost frames. Hence, an object of this paper is to reduce consumed wireless resource due to packet retransmission by improving transmission performance through diversity effect and error-correcting codes. In the proposed system, improvement is made possible especially for transmission errors of wireless terminals, which exist in cell-edge areas. The proposed mechanisms employ error-correcting codes for some frames, which consist of an IP datagram. As the results, receivers need not to wait for receptions of some IP datagrams, and only wait for receptions of all frames, which consist of an IP datagram. In numerical results, we evaluate performance by using computer simulations, and the results show that reliable multicast delivery can be performed with a few wireless resources by using the proposed system. Moreover, the proposed system can achieve multicast communication with short delivery delay.

## II. Principles of the proposed system

In this paper, we evaluate the multicast delivery system which employs base station diversity and error-correcting codes. The proposed system consists of the base station controller, some base stations, and many wireless terminals. Source terminals are assumed to exist in the Internet. Therefore, all internet protocol (IP) datagrams arrive at the base station controller from the Internet. As the results, the base station controller can manage all multicast communication to wireless terminals in the proposed system.

Figure 1 shows the proposed system that employs the base station diversity and error-correcting codes. An IP datagram is fragmented to many wireless frames in wireless systems, when we aim reliable multicast over IP networks. Wireless terminals should receive all wireless frames, which consist of an IP datagram, to serialize the fragmented wireless frame. In the general wireless systems, error-correcting codes are applied to each frame. Therefore, the proposed system can also apply error-correcting codes to each frame to improve frame error ratio. Additionally, our system employs error-correcting codes for through the all frames, which consist of an IP datagram, to improve packet error ratio. Hereinafter, we focus on the error-correcting codes to improve packet error ratio. In the proposes system, the error-correcting code generates $k$ original frames, which consist of an IP datagram, and $n - k$ error-correcting frames. Wireless terminals can reconstruct an IP datagram when they can receive $k$ frames among $n$ generated frames because the error-correcting code is applied to through the all frames.

Similar techniques have been proposed in wired networks[1]. Meanwhile, conventional mechanisms employ error-correcting codes for many IP datagrams. Therefore, receivers should wait for receptions of all IP datagrams which are encoded by error-correcting codes. As the results, the increase of waiting period increases delivery delay. Additionally, performance with base station diversity and error-correcting codes has not been considered in conventional works. On the contrary, the proposed mechanisms employ error-correcting codes for some frames, which consist of an IP datagram. Therefore, receivers need not to wait for receptions of some IP datagrams, and wait for receptions of all frames, which consist of an IP datagram. Additionally, the number of retransmission frames can be reduced because base stations transmit a redundant frame to recover lost frames. As the results, the proposed mechanisms can obtain improvement of packet delivery without increasing of delivery delay.

## III. Operations in proposed system

In this paper, we assume that wireless terminals join a multicast group, and a source terminal on the Internet transmits data packets to the multicast group. Therefore, all data packets arrive at the base station controller, and they are delivered to wireless terminals via downlinks. Additionally, negative acknowledgement (NACK) is replied to base stations via uplinks by wireless terminals.

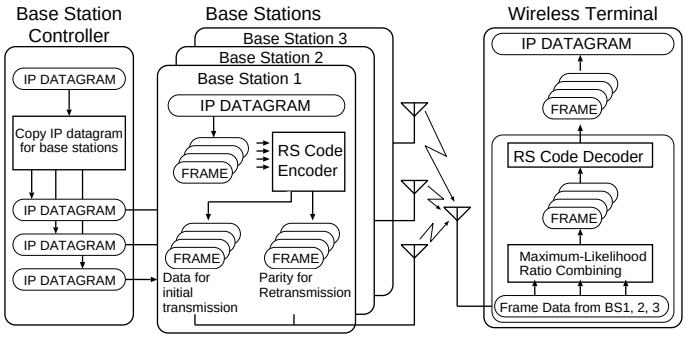



Fig. 1. System model.

### A. Data delivery to wireless terminals

- Selection of base station by wireless terminals
  Wireless terminals measure received signal intensity for neighbor base stations. They determine to communicate with some base stations when the highest signal intensity is lower than a threshold. The selected base stations are registered to the base station controller via uplink communication.
- Duplication and forwarding of IP datagrams by base station controller
  When the base station controller receives IP datagrams from a source terminal on the Internet to a multicast group, it duplicates the IP datagrams for base stations and forwards the duplicated IP datagrams to the base stations.
- Generation of redundant frames and transmission of frames by base stations
  When the base stations receive the IP datagrams from the base station controller, they divide the IP datagram into some wireless frames, which are 37 frames in simulations. Then, they generate redundant frames for retransmission due to frame errors by error-correcting codes. As the results, the base stations store $k$ original frames which consist of the IP datagram, and $n - k$ redundant frames for retransmission. Finally, they start to transmit $k$ original frames for initial transmission.
- Frame reception by wireless terminals
  The wireless terminals wait for frame receiving from neighbor base stations. They estimate SNR of the received frames from neighbor base stations and temporarily store soft decision values of each bit. When the frame receiving from all of the desired base stations is completed, the frame data are decoded by MRC (Maximal Ratio Combining) based on stored SNRs of the frames and soft decision values. As the results, the wireless terminals can obtain diversity effect through the neighbor base stations, and improve packet error ratio.

### B. Operations of retransmission for lost frames

- Request for frame retransmission by wireless terminals
  The wireless terminals check received frame data. Since they need $k$ frames, which are original or redundant frames, to decode an IP datagram by error-correcting codes, they count the number of received frame data. Then, wireless terminals request frame retransmission to the base stations by transmitting NACK frames that include the number of additional number of frame data if they detect frame data errors.
- Transmission of redundant frames by base stations
  After frame transmission to the wireless terminals, the base stations wait for NACK frames, which request frame retransmission. After receptions of the NACK frames from the wireless terminals, the base stations tally up requested frames by checking the number of additional number of frame data in the NACKs for retransmission process. Then, they count the maximum number of frame losses $r$. Finally, they transmit $r$ redundant frames, which is the same number of maximum lost frames. These procedures are performed until the number of retransmissions becomes the maximum retransmission number or the all $n-k$ prepared redundant frames are retransmitted.
- Frame decoding by wireless terminals
  The wireless terminals can decode the IP datagram when the number of received frames becomes more than the number of original frames, which equals to $k$. Finally, they can utilize redundant frames by using error-correcting codes to improver performance of IP datagram delivery. These procedures are performed for each IP datagram. Therefore, the delivery delay does not increase by employing the proposed system.

## IV. Numerical results

In this section, we evaluate the performance of reliable multicast delivery on the proposed system by using computer simulations. The results are calculated from averages of 100000 wireless terminal's results and the number of wireless terminals is set to 10, 100, 1000 and 10000, and all wireless terminals are located randomly within 18 cell's area in the simulations. In the simulations, each wireless terminal joins the same multicast group. The source terminal exists on the wired network, and transmits CBR traffic to the multicast group. Therefore, almost all data traffic is transferred over

TABLE I

SIMULATION PARAMETERS

| | |
|---|---|
| Transmitting power | 90[dB] at 1[m] from BS |
| Path loss exponent | 3.5 |
| Channel | Rayleigh fading |
| Modulation scheme | BPSK |
| Bandwidth | 11[Mbps] |
| Number of cell | 18 |
| Cell radius | 250[m] |
| Number of channels | 4 |
| Frame size | 127 [bit] |
| FEC | Reed Solomon Code (255, 37) |
| Datagram size | 576[Byte] |
| Wired network latency | 25[ms] |
| Application | CBR 512[kbps] |
| Wireless terminal distribution | Uniform |
| Number of user | 10 - 10000 |
| Mobility | None |
| Diversity combining | MRC |
| Number of connected base stations | 3 |
| Retransmission limit | 9 times |

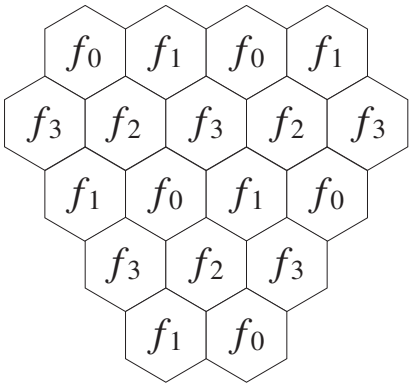



Fig. 2. Channel allocation.

the downlink, and only NACK traffic is transferred over the uplink.

Figure 2 shows channel allocation in the simulations. In the simulations, we assume one of the four channels is allocated to each cell. Meanwhile, two adjacent cells do not share the same channel. Hence, we consider the inter-cell interference between same frequency channel allocated not adjacently in the simulations. Additionally, each base station has two types of channel resources that are used for uplink and downlink communications. Therefore, both uplink and downlink communications perform simultaneously without interference. Additionally, all wireless terminals in each cell share the same channel. Therefore, the channel access is controlled by each base station. In order to evaluate fundamental performance, mobility of wireless terminals is not considered. However, the proposed system can apply to mobility environment. In the simulations, we consider the three type conventional systems, which are cellular systems with only single link, cellular systems with single link and error-correcting codes, and cellular system with multiple links. Detail simulation parameters are shown in Table I.

Figure 3 shows the frame transmission ratio versus the number of retransmissions. The frame retransmission ratio means the ratio of retransmitted frames per 37 frames, which consist of one IP datagram. Therefore, 1 means that base stations transmit 37 frames, which consist of one IP datagram. The number of retransmission means times of retransmission. Therefore, 0 means an initial transmission process and 1 means a first retransmission process. From the results, base stations of the conventional system retransmit many wireless frames repetitively. Additionally, those of the base station diversity system also retransmit many wireless frames repetitively. On the contrary, those with error-correcting codes retransmit a few wireless frames because only redundant frames are transmitted for lost frames. From the results, error-correcting codes are effective to reduce the number of transmitted frames in cellular systems.

Figure 4 shows the reception ratio of IP datagrams. From the results, we can confirm that eight times of retransmission are required to deliver IP datagrams to all wireless terminals in the conventional system. Moreover, more than nine times of retransmission are required in the conventional system with error-correcting codes. On the contrary, we can find that less than four times of retransmission are enough to deliver IP datagrams to all wireless terminals in the base station diversity systems. As the results, the proposed system can improve the reception ratio of IP datagrams with a fewer numbers of transmitted frames. Additionally, the reduction of the retransmission times means the reduction of delivery delay. Therefore, the proposed system also effective to reduce delivery delay in reliable multicasting.

Figure 5 shows the ratio of the total frame transmission versus the number of wireless terminals. It is a value that total transmitted wireless frames are normalized by 37 frames, which consist of one IP datagram. Therefore, the value is one if the error-free wireless link can be assumed. From the results, we can confirm that required wireless resource of the conventional system increases with increasing numbers of wireless terminals because the number of requested frames increases when many wireless terminals communicate. In contrast, the number of transmitted frames can be reduced in

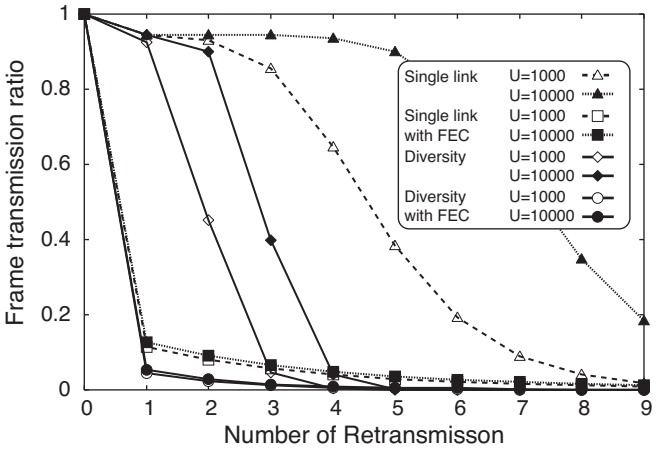



Fig. 3. Frame transmission ratio.

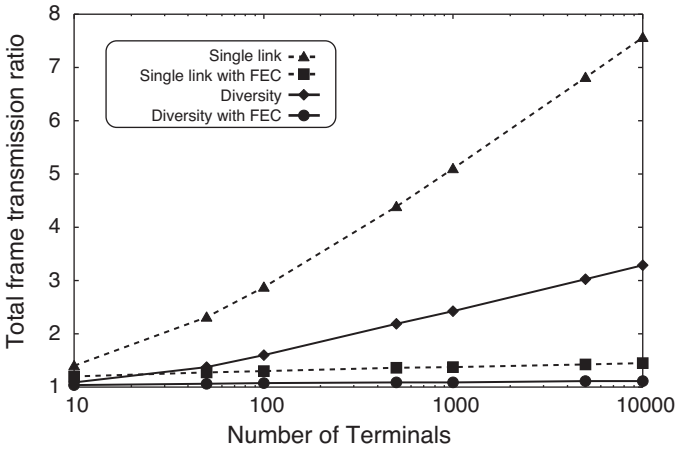



Fig. 5. Total frame transmission ratio.

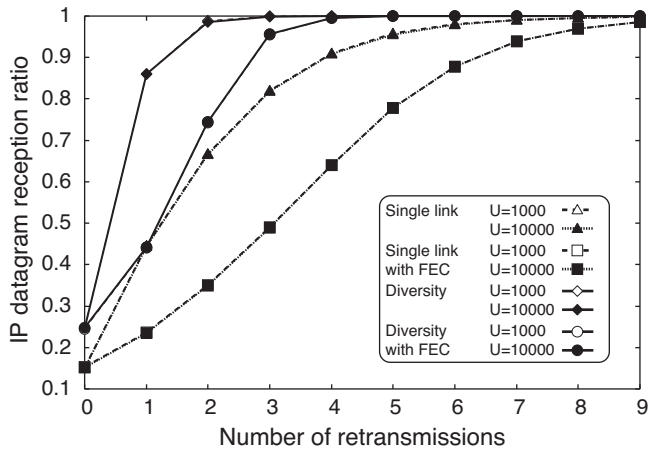



Fig. 4. IP datagram reception ratio.

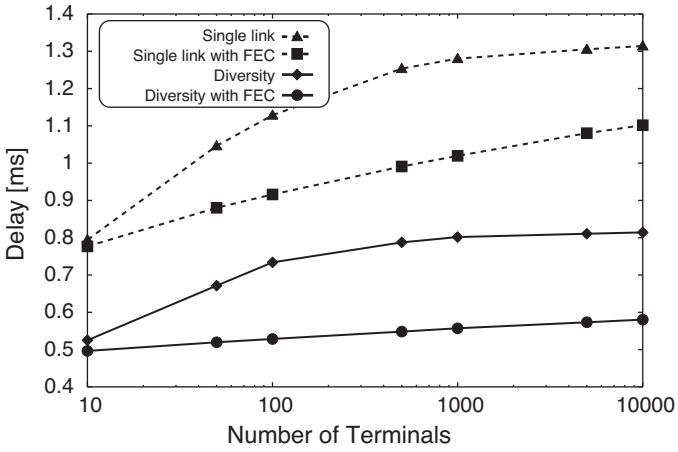



Fig. 6. Delay performance.

the proposed system because almost all wireless frames can be received at the initial transmission. Moreover, redundant frames are common retransmission frames for lost frames. Therefore, the number of frame transmission can be reduced.

Figure 6 shows the delay performance versus the number of wireless terminals. The delay is the required period to achieve 100% delivery of the IP datagram. Therefore, the delay period is about 0.43[ms] according to the simulation parameters if there is no transmission error. From the results, we can find that the delay of the systems without error-correcting codes increases. On the contrary, we can confirm that the delay of the systems with error-correcting codes can reduce the delay period. Moreover, the base station diversity with error-correcting codes can keep the short period even if the number of wireless terminals increases.

## V. Conclusions

In this paper, we proposed the new reliable multicast delivery system which employs base station diversity techniques and error-correcting codes. In the proposed system, wireless terminals receive some wireless signals from neighbor different base stations, and decode received frames by MRC method. Additionally, base stations generate redundant frames by error-correcting codes for retransmission due to frame errors. Therefore, base stations can transmit redundant frames in retransmission for lost frames. The numerical results showed that the proposed system maintained the better performance compared to the conventional cellular system. Moreover, the proposed system can reduce the occupancy of wireless resource by improving transmission performance by base station diversity and reduction of retransmission frames by error-correcting codes.

## ACKNOWLEDGMENT

This work was supported by the Telecommunications Advancement Foundation and MEXT KAKENHI(23700075).

## REFERENCES

[1] S. Paul, K. K. Sabnani, J. C. Lin and S. Bhattacharyya, "Reliable Multicast Transport Protocol (RMTP)," IEEE Journal on Selected Areas in Communications, vol. 15, no. 3, pp. 407–421, Apr. 1997.

[2] L. Rizzo, "Effective Erasure Codes for Reliable Computer Communication Protocols," ACM SIGCOMM Computer Communication Review, vol. 27, pp. 24–36, Apr. 1997.

[3] R. Rummler, Yun Won Chung, A.H. Aghvami, "Modeling and analysis of an efficient multicast mechanism for UMTS," IEEE Transactions on Vehicular Technology, Vol. 54, No. 1, pp. 350–365, Jan. 2005.

[4] De-Nian Yang, Ming-Syan Chen, "Efficient Resource Allocation for Wireless Multicast," IEEE Transactions on Mobile Computing, Vol. 7, No. 4, pp. 387–400, Apr. 2008.

[5] Seung Young Park, David J. Love, "Outage performance of multi-antenna multicasting for wireless networks, " IEEE Transactions on Wireless Communications, Vol. 8, No. 4, pp. 1996–2005, Apr. 2009.

[6] K. Sundaresan, K. Ramachandran and S Rangarajan, "Optimal Beam Scheduling for Multicasting in Wireless Networks," ACM MobiCom '09, Sep. 2009.

[7] A. Alexiou, C. Bouras and A. Papazois, "An Efficient Mechanism for UMTS Multicast Routing," Journal Mobile Networks and Applications, vol. 15, no. 6, Dec. 2010.

[8] Suttipong Thajchayapong, Jon M. Peha, "Mobility Patterns in Microcellular Wireless Networks, " IEEE Transactions on Mobile Computing, Vol. 5, No. 1, pp. 52–63, Jan. 2006.

[9] S. M. Alamouti "A Simple Transmit Diversity Technique for Wireless Communications," IEEE Journal on Select Areas in Communications, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[10] A. Miu, H. Balakrishnan and C. E. Koksal, "Multi-radio diversity in wireless networks," Journal Wireless Networks, vol. 13, no. 6, Dec. 2007.

[11] H. Yu and G. L. Stüber, "Outage probability for decode-and-forward cooperative diversity with selective combining in cellular networks," ACM IWCMC '10, Jun. 2010.

[12] Katsuhiro NAITO, Kazuo MORI and Hideo KOBAYASHI, "Evaluation of reliable multicast delivery in base station diversity systems," IEICE Transactions on Communications, Vol. E93-B, No. 6, pp. 1615–1619, Jun. 2010.

# Failure estimation technique for optical access networks based on ONU information and topology information

Katsuhiro Naito, Kazuo Mori, and Hideo Kobayashi
Department of Electrical and Electronic Engineering, Mie University,
1577 Kurimamachiya, Tsu, 514-8507, Japan
Email: {naito, kmori, koba}@elec.mie-u.ac.jp

Makoto Nishide
NetStep Co., Ltd.,
231 ObatachoMiyamae, Ise, 519-0504, Japan
Email: mnishide@netstep.co.jp

***Abstract*— Passive optical networks (PONs) are drawing attention to provide high-speed network services. PONs consist of only passive elements such as optical fibers and splitters. Some optical network units (ONUs) share a single optical fiber to communicate with an optical line terminal (OLT). In the conventional works, some monitoring schemes for optical fibers have been proposed. Optical time domain reflectometers (OTDRs) are special equipment to sense optical fibers by using reflect waves. Therefore, a lot of studies employ OTDRs to monitor the optical fibers. However, OTDRs suffer from lack of dynamic range after splitter in tree-structured PONs. Additionally, implementation cost of OTDRs is generally expensive to be installed into large size networks.**

**In this paper, we propose a new failure estimation technique for optical access networks based on ONU information and topology information. The feature of the proposed technique is to utilize information about optical fibers collected by ONUs. Some ONUs connecting to the same optical fiber detect link down when link failures occur over the optical fiber. Therefore, the system tallies up the number of ONUs that detect link failures at each device by referring to the topology information. Then, it estimates devices that have relations with the link failures.**

***Keywords*— Optical fiber, Passive optical networks, Monitoring, Failure estimation**

## I. Introduction

Passive optical networks (PONs) are new network technologies providing fiber-to-the-home (FTTH) services. Only passive elements between an optical line terminal (OLT) and some optical network units (ONUs) are installed in PONs. ONUs share an optical fiber and communicate with an OLT [1], [2], [3].

In real devices of ONUs and OLTs, they support to check their devices by self maintenance mechanisms. Therefore, device failures can be detected by special management software provided by device makers. On the contrary, failures over optical fibers are not maintained by the management software in general systems for PONs. It is clear that reliable monitoring systems are required for practical implementation of PONs[4]. As the results, monitoring technologies for PONs have been proposed [5], [6].

Several physical layer monitoring schemes using optical time domain reflectometers (OTDRs) have been proposed [7], [8], [9], [10]. OTDRs can sense optical fibers by using reflect waves. Therefore, it can estimate a position of failures over optical fibers. However, OTDRs suffer from lack of dynamic range after splitter in tree-structured PONs. Additionally, implementation cost of OTDRs is generally expensive to be installed into large size networks.

Meanwhile, a computer aided design (CAD) has been used to design practical PONs. Therefore, topology information can be obtained in real PONs. In this paper, we propose a new failure estimation technique for optical access networks based on ONU information and topology information. The feature of the proposed technique is to utilize information about optical fibers collected by ONUs. Some ONUs connecting to the same optical fiber detect link down when failures occur over the optical fiber. Additionally, many optical fibers are installed as an optical fiber cable containing many optical fibers and a wire rope. Therefore, damages to optical fiber cables cause some failures over some optical fibers. As the results, some ONUs connecting to the optical fiber cables detect link down at almost same instance. In the proposed technique, the system tallies up the number of ONUs that detect link down at each device by referring to the topology information. Then, it estimates devices related to the link failures.

## II. Principles of the proposed system

Figure 1 shows the system model of the proposed technique. In the proposed system, the topology information of PONs is assumed to be obtained from the CAD information and Subscriber Management System (SMS) information beforehand. In the assumed PONs, ONUs can detect link down of their optical fibers. Additionally, OLTs can collect the information

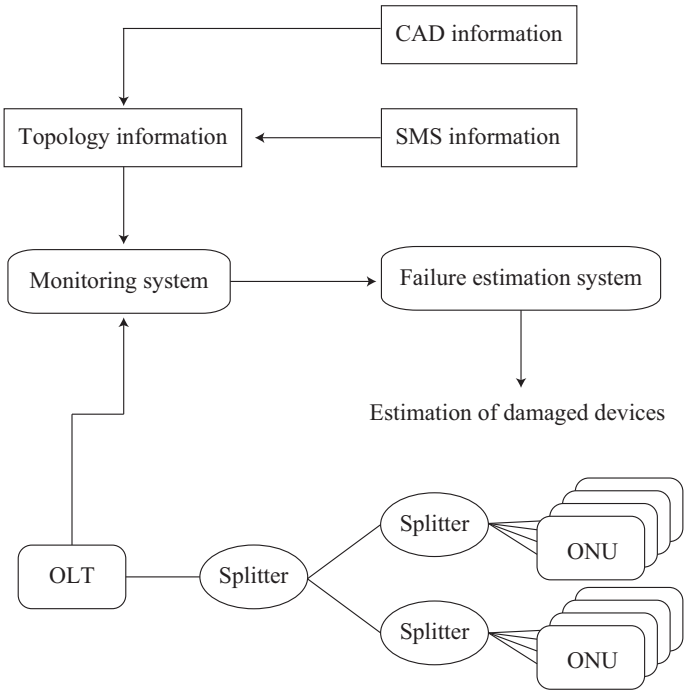



Fig. 1. System model.

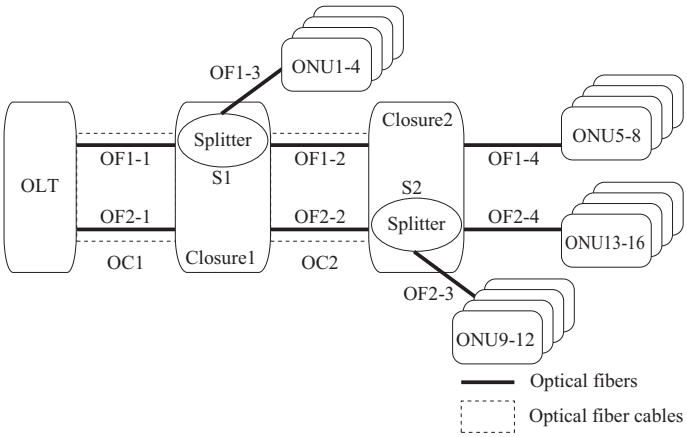



Fig. 2. Assumed optical networks in principle examples.

about link down by using vender implemented mechanisms. Finally, the proposed failure estimation system can obtain the link down information of optical fibers, that each ONU detected. Then, the failure estimation system tallies up the number of ONUs that detect link down for each device and optical fiber. As the results, the system can estimate positions related to the link down of optical fibers.

Figure 2 shows the assumed optical network in principle examples. In the example network, the two optical fibers are connected to the OLT. These optical fibers are contained by the optical fiber cable. Optical fiber cables are connected at

TABLE I

EXAMPLE NUMBER OF ONUS AT DEVICES.

| Device | Number of ONUs | ONUs |
|---|---|---|
| OF1-1 | 8 | ONU1-8 |
| OF1-2 | 4 | ONU5-8 |
| OF1-3 | 4 | ONU1-4 |
| OF1-4 | 4 | ONU5-8 |
| OF2-1 | 8 | ONU9-16 |
| OF2-2 | 8 | ONU9-16 |
| OF2-3 | 4 | ONU9-12 |
| OF2-4 | 4 | ONU13-16 |
| S1 | 8 | ONU1-8 |
| S2 | 8 | ONU9-16 |
| OC1 | 16 | ONU1-16 |
| OC2 | 12 | ONU5-16 |

closures. Additionally, Splitters for PONs are equipped in closures. Therefore, the optical fiber OF1-1 is connected to the OLT and the splitter S1, and the optical fibers OF1-2 and OF1-3 are connected to the splitter S1. Finally, the ONUs 1, 2, 3 and 4 connect to the optical fiber OF1-3, and the ONUs 5, 6, 7 and 8 connect to the optical fiber OF1-4. In a similar fashion, the ONUs 9, 10, 11 and 12 connect to the optical fiber OF2-3, and the ONUs 13, 14, 15 and 16 connect to the optical fiber OF2-4.

Table I is the number of ONUs at the devices in Fig. 2. From the table, we can obtain the number of ONUs through each device. For example, the number of ONUs through the optical fiber OF1-1 is 8 because the ONU 1-8 are through the optical fiber OF1-1. In this paper, we can obtain the number of ONUs by checking topology information from CAD systems beforehand.

Figure 3 shows the principle example with a damaged optical fiber. In the example, the optical fiber OF1-2 has some damages by extrinsic compression. As the results, the ONUs from 5 to 8 detect link down of their optical fibers. In PONs, failure information about link down is informed to an OLT. Therefore, the proposed monitoring system can obtain the failure information from the ONUs from 5 to 8. Finally, the failure estimation system confirms that the ONUs from 1 to 4 and from 9 to 16 are available in service and the ONUs from 5 to 8 are unavailable in service. Then, it calculates the link down ONUs in the similar fashion of Tab. I. As the results, it can obtain the number of link down ONUs in Tab.II. In Tab. II, it can obtain the link down ratio by checking the number of ONUs and the number of link down ONUs. Then, it can detect devices containing all link down ONUs. In this example, OF1-1, OF1-2, OF1-4, S1, OC1 and OC2 contain all link down ONUs. Therefore, it estimates that these devices are the candidate damaged devices. From the results,

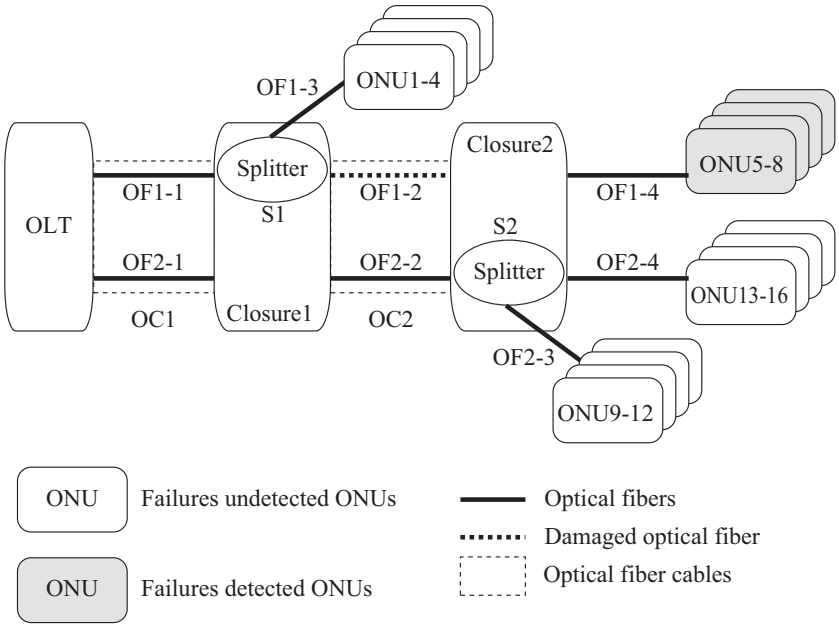



Fig. 3. Principle example with the damaged optical fiber.

TABLE II

LINK DOWN RATIO WITH THE DAMAGED OPTICAL FIBER.

| Device | Number of ONUs | Number of link down ONUs | Link down ratio | Link down ONUs | Containing all link down ONUs | Estimated damaged devices |
|---|---|---|---|---|---|---|
| OF1-1 | 8 | 4 | 0.5 | ONU5-8 | * | |
| OF1-2 | 4 | 4 | 1 | ONU5-8 | * | * |
| OF1-3 | 4 | 0 | 0 | None | | |
| OF1-4 | 4 | 4 | 1 | ONU5-8 | * | * |
| OF2-1 | 8 | 0 | 0 | None | | |
| OF2-2 | 8 | 0 | 0 | None | | |
| OF2-3 | 4 | 0 | 0 | None | | |
| OF2-4 | 4 | 0 | 0 | None | | |
| S1 | 8 | 4 | 0.5 | ONU5-8 | * | |
| S2 | 8 | 0 | 0 | None | | |
| OC1 | 16 | 4 | 0.25 | ONU5-8 | * | |
| OC2 | 12 | 4 | 0.33 | ONU5-8 | * | |

the system confirms that the link down ratio of OF1-2 and OF1-4 equals to 1. Therefore, the system can detect that all ONUs through the OF1-2 and the OF1-4 are unavailable in service. As a consequence, the optical fibers OF1-2 and OF1-4 are the final candidate damaged devices according to the collected link down information. Finally, the system informs the estimated damaged devices as the optical fibers OF1-2 and OF1-4.

Figure 4 shows the principle example with a damaged optical fiber cable. In the example, the optical fiber cable OC2 has some damages by extrinsic compression and the optical fibers OF1-2 and OF2-2 are also damaged. As the results, the ONUs from 5 to 16 detect link down of the optical fibers. The failure information about the link down is informed to the monitoring system. Finally, the failure estimation system confirms that the ONUs from 1 to 4 are available in service and the ONUs from 5 to 16 are unavailable in service. Then, it calculates the link down ONUs in the similar fashion of Tab. I. As the results, it can obtain the number of link down ONUs in Tab.III. From the results, the link down ratio of OF1-2, OF1-4, OF2-1, OF2-2, OF2-3, OF2-4, S2 and OC2 equal to 1. Therefore, the system can detect that all ONUs through these devices are unavailable in service. Then, it can detect devices containing all link down ONUs. In this example, OC1 and OC2 contain all link down ONUs. Therefore, it estimates that these devices are the candidate damaged devices. Hence, it can detect that the optical fiber cable OC2 has the 100% link down ratio. Finally, it informs the estimated damaged devices

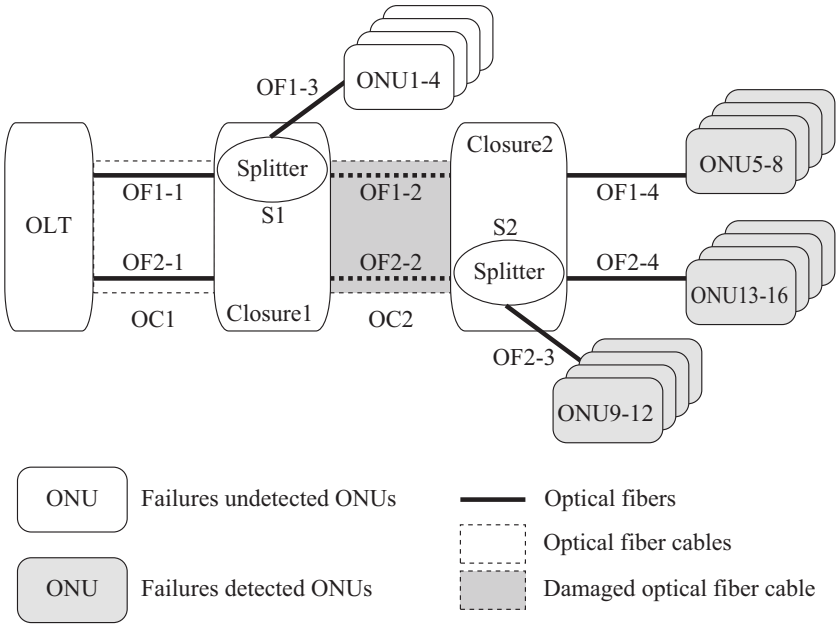



Fig. 4. Principle example with the damaged optical fiber cable.

TABLE III

LINK DOWN RATIO WITH THE DAMAGED OPTICAL FIBER CABLE.

| Device | Number of ONUs | Number of link down ONUs | Link down ratio | Link down ONUs | Containing all link down ONUs | Estimated damaged devices |
|---|---|---|---|---|---|---|
| OF1-1 | 8 | 4 | 0.5 | ONU5-8 | | |
| OF1-2 | 4 | 4 | 1 | ONU5-8 | | |
| OF1-3 | 4 | 0 | 0 | None | | |
| OF1-4 | 4 | 4 | 1 | ONU5-8 | | |
| OF2-1 | 8 | 8 | 1 | ONU9-16 | | |
| OF2-2 | 8 | 8 | 1 | ONU9-16 | | |
| OF2-3 | 4 | 4 | 1 | ONU9-12 | | |
| OF2-4 | 4 | 4 | 1 | ONU13-16 | | |
| S1 | 8 | 4 | 0.5 | ONU5-8 | | |
| S2 | 8 | 8 | 1 | ONU9-16 | | |
| OC1 | 16 | 12 | 0.75 | ONU5-16 | * | |
| OC2 | 12 | 12 | 1 | ONU5-16 | * | * |

as the optical fiber cable OC2.

## III. ESTIMATION OF FAILURE POSITION

Figure 5 shows the flowchart of the proposed system. In order to estimate failure position in PONs, the following procedures are performed.

- Initialization of the topology information in PONs
  The system tallies up the number of ONUs through each device between OLT and ONUs by checking the CAD information and SMS information. As the results, the system can obtain the list that includes the total number of ONUs at each device.
- Detection of link failures
  ONUs detect link down when some failures occur over the PONs. The detected link down information is informed to the monitoring system for PONs by using vendor maintenance mechanisms. For example, notification methods are assumed to be simple network management protocol (SNMP). item Aggregation of link failure information In order to separate reasons of the link failures, the system aggregates the link failure information by using detected time stamp information. As the results, link failure information detected nearly simultaneously is aggregated to one failure event.
- Tallying up the number of link down ONUs
  In order to evaluate relationship of each device to the detected link down event, the system tallies up the number of link down ONUs through each device in PONs.

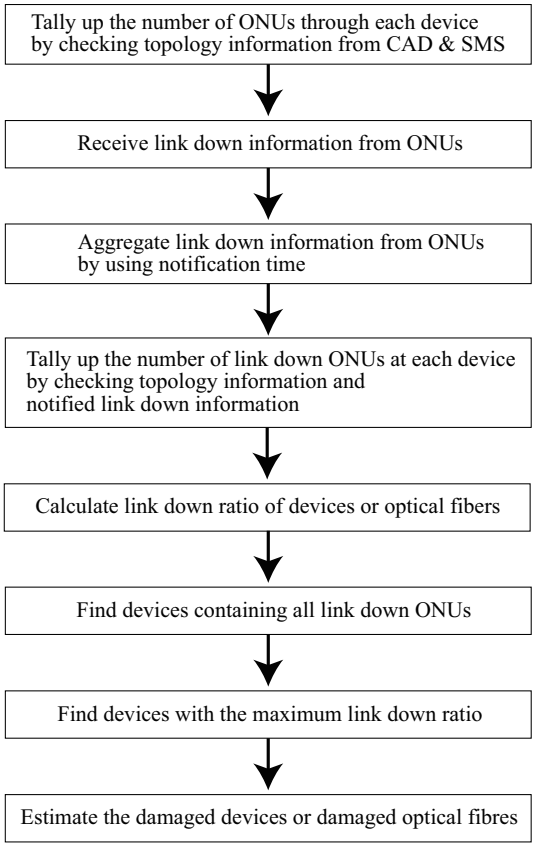



Fig. 5. Flowchart of the proposed system.

- Calculation of link down ratio
  In this paper, we define a link down ratio, in which the number of link down ONUs is divided by the number of total ONUs for each device. The link down ratio with 1 means that all ONUs through the device detect link down.
- Find devices containing all link down ONUs
  In order to find devices that effect all link down ONUs, the system find devices containing all link down ONUs. As the results, the system estimates the devices that have some relations to all link down ONUs.
- Find devices with the maximum link down ratio
  In order to detect devices which are a main cause for link down at all link down ONUs, the system picks up devices with the maximum link down ratio.
  As the results, it estimates the damaged devices that cause the link down at all link down ONUs.

## IV. Example estimation

Figure 6 shows the assumed example network. The four optical fibers are connected to the OLT. These optical fibers are contained by the optical fiber cables. Additionally, some optical fibers are aggregated as the tapes in the optical fiber cables. Each optical fiber cable is connected at closures. Splitters for PONs are equipped in closures. Each optical fiber is connected to eight ONUs.

In the example, the tape T2-1 in the optical fiber cable OC2 has damages from extrinsic compression. Generally, some optical fibers are bonded as tapes in optical fiber cables. Therefore, damages to the tape cause failures of optical fibers in the damaged tape. As the results, the ONUs from 5 to 16 detect link down in the example.

Table IV shows the summary of the total number of connected ONUs, the total number of link down ONUs, the calculated link down ratio, the ONUs detecting link down, the devices containing all link down ONUs and the estimated damaged devices. From the summary, the system can find that T1-1, T2-1, OC1 and OC2 contain the all link down ONUs. Additionally, it picks up the maximum link down ratio among these devices. As the results, it estimates that the tape T2-1 is the candidate damaged device for link down at ONU5-16.

Figure 7 shows the assumed example network. In the example, the tape T1-1 in the optical fiber cable OC1 has damages from extrinsic compression. In this example, we consider the partial damages in tapes. Therefore, the partial damages in the tape T1-1 cause the link down at ONU1-16. On the contrary, ONU17-24 are available in service even if they through the tape T1-1.

Table IV shows the summary of the total number of connected ONUs, the total number of link down ONUs, the calculated link down ratio, the ONUs detecting link down, the devices containing all link down ONUs and the estimated damaged devices. From the summary, the system can find that T1-1, and OC1 contain the all link down ONUs. Additionally, it picks up the maximum link down ratio among these devices. As the results, it estimate that the tape T1-1 is the candidate damaged device for link down at ONU1-16.

## V. Conclusions

In this paper, we proposed the new failure estimation technique for optical networks. The proposed technique utilizes information about optical fibers monitored by ONUs and topology information obtained from CAD systems. The proposed system can estimate locations or devices that cause link failures by tallying up a number of total ONUs through each device and a number of total ONUs that detect link down. As the results, the proposed system can achieve the novel monitoring scheme without any additional equipment such as OTDRs into PONs.

## Acknowledgment

This work was supported by NetStep Co., Ltd.

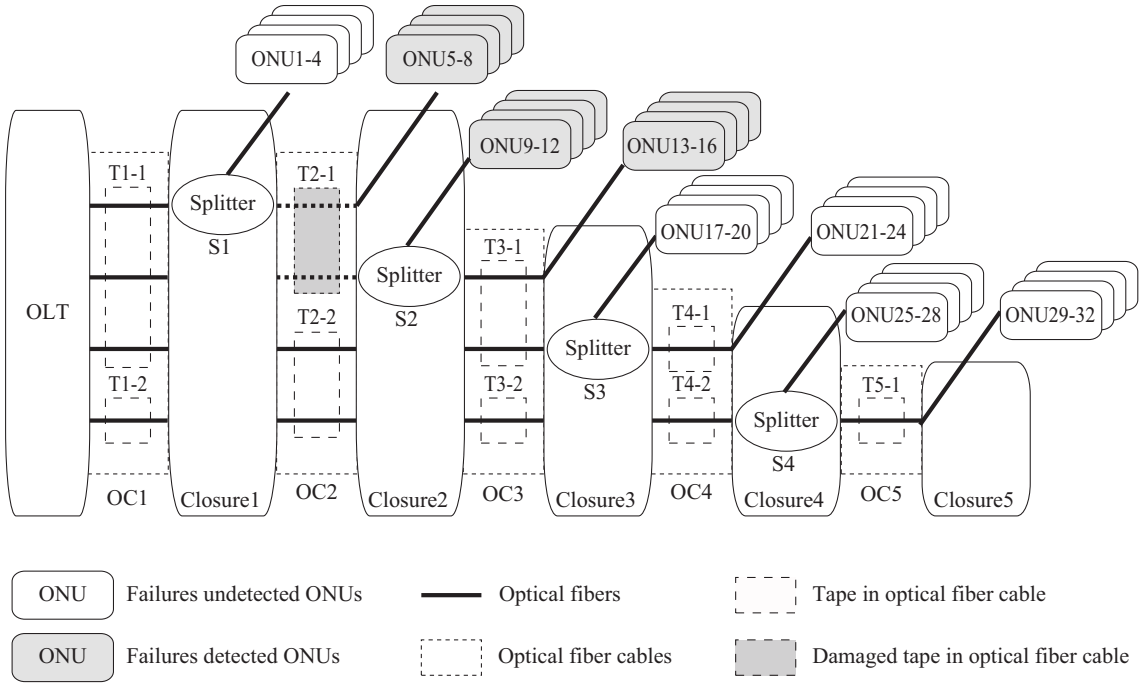



Fig. 6. Example network with the damaged tape.

TABLE IV

LINK DOWN RATIO WITH THE DAMAGED TAPE.

| Device | Number of ONUs | ONUs | Number of link down ONUs | Link down ratio | Link down ONUs | Containing all link down ONUs | Estimated damaged devices |
|---|---|---|---|---|---|---|---|
| T1-1 | 24 | ONU1-24 | 12 | 0.5 | ONU5-16 | * | |
| T1-2 | 8 | ONU25-32 | 0 | 0 | | | |
| T2-1 | 12 | ONU5-16 | 12 | 1 | ONU5-16 | * | * |
| T2-2 | 16 | ONU17-32 | 0 | 0 | | | |
| T3-1 | 12 | ONU13-24 | 4 | 0.33 | ONU13-16 | | |
| T3-2 | 8 | ONU25-32 | 0 | 0 | | | |
| T4-1 | 4 | ONU21-24 | 0 | 0 | | | |
| T4-2 | 8 | ONU25-32 | 0 | 0 | | | |
| T5-1 | 4 | ONU29-32 | 0 | 0 | | | |
| S1 | 8 | ONU1-8 | 4 | 0.5 | ONU5-8 | | |
| S2 | 8 | ONU9-16 | 8 | 1 | ONU9-16 | | |
| S3 | 8 | ONU17-24 | 0 | 0 | | | |
| S4 | 8 | ONU25-32 | 0 | 0 | | | |
| OC1 | 32 | ONU1-32 | 12 | 0.375 | ONU5-16 | * | |
| OC2 | 28 | ONU5-32 | 12 | 0.429 | ONU5-16 | * | |
| OC3 | 20 | ONU13-32 | 4 | 0.2 | ONU13-16 | | |
| OC4 | 12 | ONU21-32 | 0 | 0 | | | |
| OC5 | 4 | ONU29-32 | 0 | 0 | | | |

## REFERENCES


[1] A. Girard, "FTTx PON technology and testing," in EXFO Electro Optical Engineering Inc., Canada, 2005.

[2] X. Bai, A. Shami, and C. Assi, "On the fairness of dynamic bandwidth allocation schemes in Ethernet passive optical networks," Computer Communications Vol. 29 No. 11, pp. 2123–2135, Jul. 2006.

[3] F. Effenberger, D. Clearly, O. Haran, G. Kramer, R. D. Li, M. Oron, and T. Pfeiffer, "An Introduction to PON Technologies," IEEE Communications Magazine, vol. 45, no. 3, pp. 17–25, Mar. 2007.

[4] J. Kim, S. Mun, H. Lee and C. Lee "Self-Restorable WDM-PON With a Color-Free Optical Source," IEEE Journal of Optical Communications and Network, vol. 1, no. 6, pp. 565–570, Nov. 2009.

[5] T. Pfeiffer, J. Hehmann, H. Schmuck, W. Freude, J. Vandewege and Hiroki Yanagisawa, "Monitoring and protecting the optical layer in FTTH networks," in Proceedings of the 2005 FTTH Conference, 2005.

[6] S. Hann, J. S. Yoo, and C. S. Park, "Monitoring technique for a hybrid PS/WDM-PON by using a tunable OTDR and FBGs," Measurement Science and Technology vol. 17 No. 5, pp. 1070–1074, 2006.

[7] F.J. Effenberger, and S. Meng, "In-band optical frequency domain

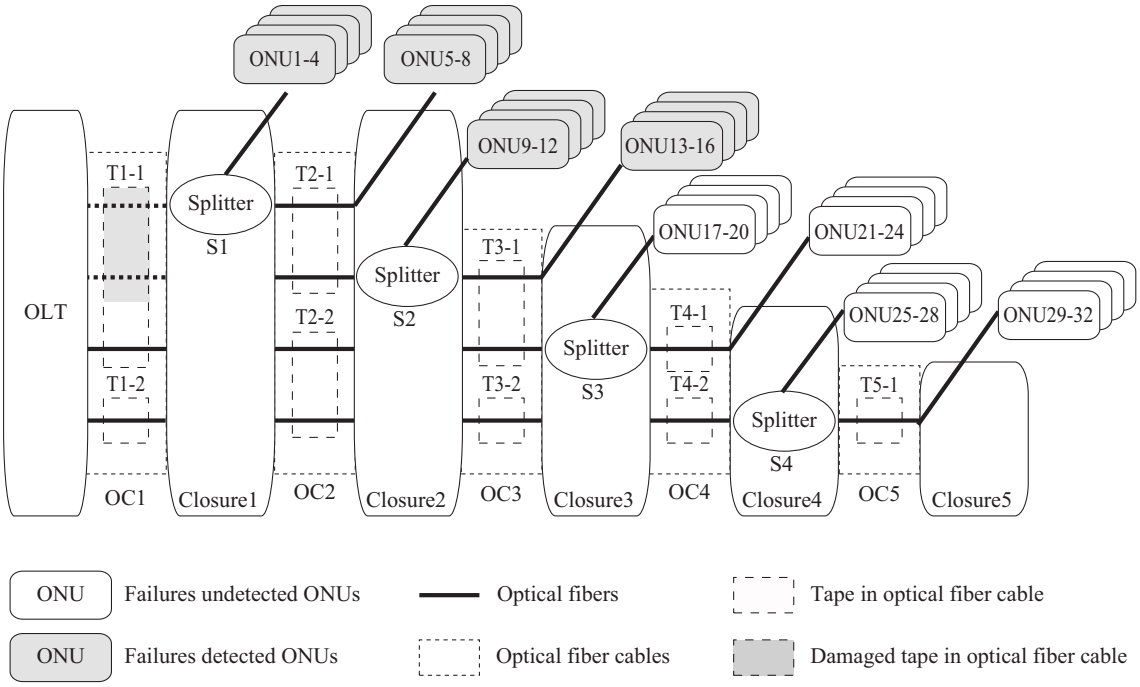



Fig. 7. Example network with the partial damaged tape.

TABLE V

LINK DOWN RATIO WITH THE PARTIAL DAMAGED TAPE.

| Device | Number of ONUs | ONUs | Number of link down ONUs | Link down ratio | Link down ONUs | Containing all link down ONUs | Estimated damaged devices |
|---|---|---|---|---|---|---|---|
| T1-1 | 24 | ONU1-24 | 16 | 0.66 | ONU1-16 | * | * |
| T1-2 | 8 | ONU25-32 | 0 | 0 | | | |
| T2-1 | 12 | ONU5-16 | 12 | 1 | ONU5-16 | | |
| T2-2 | 16 | ONU17-32 | 0 | 0 | | | |
| T3-1 | 12 | ONU13-24 | 0 | 0 | | | |
| T3-2 | 8 | ONU25-32 | 0 | 0 | | | |
| T4-1 | 4 | ONU21-24 | 0 | 0 | | | |
| T4-2 | 8 | ONU25-32 | 0 | 0 | | | |
| T5-1 | 4 | ONU29-32 | 0 | 0 | | | |
| S1 | 8 | ONU1-8 | 8 | 1 | ONU1-8 | | |
| S2 | 8 | ONU9-16 | 8 | 1 | ONU9-16 | | |
| S3 | 8 | ONU17-24 | 0 | 0 | | | |
| S4 | 8 | ONU25-32 | 0 | 0 | | | |
| OC1 | 32 | ONU1-32 | 16 | 0.5 | ONU1-16 | * | |
| OC2 | 28 | ONU5-32 | 16 | 0.57 | ONU9-16 | | |
| OC3 | 20 | ONU13-32 | 0 | 0 | | | |
| OC4 | 12 | ONU21-32 | 0 | 0 | | | |
| OC5 | 4 | ONU29-32 | 0 | 0 | | | |

reflectometry in PONs," in Proceedings of the 2008 Optical Fiber communication/National Fiber Optic Engineers Conference, pp. 1–3, Feb. 2008.

[8] N. Honda, D. Iida, H. Izumita, and Y. Azuma, "In-Service Line Monitoring System in PONs Using 1650-nm Brillouin OTDR and Fibers With Individually Assigned BFSs," IEEE Journal of Lightwave Technology, Vol. 27, No. 20, pp. 4575–4582, Oct. 2009.

[9] I. Ogushi, N. Araki, and Y. Azuma, "A New Optical Fiber Line Testing Function for Service Construction Work That Checks for Optical Filters Below an Optical Splitter in a PON," IEEE Journal of Lightwave Technology, Vol. 27, No. 23, pp. 5385–5393, Dec. 2009.

[10] J. H. Lee, J. Moon, K. Choi and C. Lee, "Seamless Maintenance and Protection Scheme for Next-Generation Access Networks," IEEE Photonics Technology Letters, vol. 21, no. 12, Jun. 2009.

[11] K. Yüksel, M. Wuilpart, V.Moeyaert, and P. Mügret, "Novel Monitoring Technique for Passive Optical Networks Based on Optical Frequency Domain Reflectometry and Fiber Bragg Gratings," IEEE Journal of Optical Communications and Networking, Vol. 2, No. 7, pp. 463–468, Nov. 2010.

# Development of ZigBee-based Life Recording and Management System on Mobile WiMAX Network in Azumino City

Kenko Ota*, Hideaki Matsue*, Satoru Miyazawa*, Satoshi Nanamatsu*, Yukihiro Hirata*,
Akio Hasui†, Masahiro Yamazaki†, Hiroshi Fukui‡, Matthew Harvey‡,
Hajime Miyajima§, Yoshiaki Yazawa§, Shunzo Yamashita¶, Yoshihiro Kainuma* and Takuma Yui*
*Department of Electronic Systems Engineering
Tokyo University of Science, Suwa
5000-1 Toyohira, Chino, Nagano 391-0292, JAPAN
†Azumino City
4810-1 Meisei, Misato Azumino, Nagano 399-8101, JAPAN
‡Commuture Corp.
SIA Toyosu Primesquare 9th F, 5-6-36 Toyosu, Koto-ku, Tokyo 135-8113, JAPAN
§Azumino Networks Community TV
574-5 Hotaka, Azumino, Nagano 399-8303, JAPAN
¶Division of WirelessInfo
Hitachi Ltd.
1-18-3 Sotokanda, Chiyoda-ku, Tokyo 101-8606, JAPAN

**ABSTRACT**

In this paper, firstly, a ZigBee-based sensor network is realized and evaluations are carried out in order to clarify the fundamental performance of this sensor network. The following evaluations are carried out: first one is an evaluation of the effect on communication time by the co-channel interference, and second one is an evaluation of relationship between throughput and distance. Next, the life recording and management system as an application of ZigBee-based sensor networks is developed based on the results of evaluations, and the system is introduced into the mobile WiMAX network in Azumino city in order to solve some medical problems in local cities. Finally, this paper shows both some examples of visualized physical condition data and the results of field trial test; and the effectiveness and future works of this system are clarified.

**Keywords**: Life recording and management system, Wearable sensor terminal, ZigBee, Mobile WiMAX, Data visualization.

## 1. INTRODUCTION

Recent drastic advancements in wireless technology are realizing the environment allowing us to connect to networks at any time and from anywhere. Wireless sensor network technology plays an important role as one of the technologies which realize the above-mentioned environment.

Although wireless sensor networks do not require high speed communication of sensors or control devices, the following characteristics are required to easily develop systems allocated a lot of devices; Low power, Low cost and Interconnectivity. IEEE802.15.4 (ZigBee)[1] satisfying the above characteristics has been formulated by ZigBee alliance[2]. Products complying with ZigBee are provided by various vendors[3][4].

ZigBee-based wireless networks are applied to disaster prevention, security, transportation, logistics, medical services, public welfare, information appliance and so on. In this paper, life recording and management system is introduced as one of the ZigBee applications.

In recent year, Japanese people have concerned about the lifestyle diseases, since the law concerning introduction of medical check and health direction for the prevention of lifestyle diseases came into force in April 2008. They begin to consider it important that they record and manage their own physical condition. However, to record and manage their own physical condition is a troublesome task. Here, this paper introduces the life recording and management system that physical condition is measured by sensors automatically, and the measured data are stored on servers through sensor networks. If these data can be shared with medical experts, people can not only escape from the troublesome task, but they can also correctly comprehend their own physical condition.

Firstly, following evaluations are carried out using a ZigBee-based sensor in order to clarify the fundamental performance of the sensor.

- Evaluation of the effect of the co-channel interference on communication time
- Evaluation of the throughput performance at each distance

Next, the life recording and management system is deployed in the mobile WiMAX network in Azumino city[5] in order to solve some medical problems in local cities. Finally, from some examples of visualized physical condition data and the results of field trial test, the effectiveness and future works of this system are clarified.

The rest of this paper consists of the following five sections. Section II denotes the overview of the life recording and management system. Then, in Section III, performance evaluation results are denoted. Section IV shows the network configuration of the system. Section V shows examples of visualized physical condition data. Section VI shows the results of field trial test. Finally, Section VII concludes this paper.

TABLE I
CONFIGURATION OF LIFE RECORDING AND MANAGEMENT SYSTEM.

| Parameters | Contents | Quantity |
|---|---|---|
| Wristwatch-type sensor terminal | ZigBee-based sensor<br>Hitachi Ltd. AirSense02Plus | 2 |
| Base station for ZigBee | Linux-BOX<br>1. Data receive and accumulation software<br>2. Ethernet LAN interface<br>3. USB-type receiver | 1<br><br><br>1 |
| Life recording and management system server | Linux-BOX<br>1. Life information visualization software | 1 |
| Client PC | Web browser<br>(for displaying life data) | 1 |

## 2. OVERVIEW OF THE LIFE RECORDING AND MANAGEMENT SYSTEM

The developed system can manage physical condition data recorded by "Hitachi Ltd. AirSense02Plus[6][7][8]" via TCP/IP network. The system provides the following functions.

- Measurement of temperature and amount of activity (acceleration)
- Visualization of daily life rhythm
- Share of physical condition data via TCP/IP network

The system consists of the elements shown in Table I; Table I shows the configuration of life recording and management system. A base station for ZigBee can distinguish up to 2 wristwatch-type sensors. It is a Linux-BOX customized for data receive and accumulation software. A life recording and management system server is also a Linux-BOX customized for life information visualization software.

### Wearable wristwatch-type ZigBee sensor

Wearable wristwatch-type ZigBee sensor is employed for measuring various physical condition data, because users can wear it like a wristwatch. Thus, to wear the sensor is not a burden for users.

Table II shows the specification of the wristwatch-type ZigBee sensor. Besides, Fig. 1 shows the pictorial image of the wristwatch-type ZigBee sensor. The wristwatch-type sensor equips triaxial acceleration and temperature sensors. Various physical condition data can be measured by these sensors. Measured data by the sensors are automatically stored on life recording and management system server via wireless network.

The reason why the battery life and data storage capacity depend on usage environments is that "differential compression" is employed for store of the measurement data. Thus, measured data size depends on the amount of changing of activity. Battery life become shorter by transmitting the all data stored on the internal data storage of wristwatch-type sensor.

### Visualization of physical condition data

Measured data by a wristwatch-type sensor are analyzed by the life recording and management system server, and these data can be displayed by web browsers. This system can display the amount of activity for a day, a week and three months. Examples of visualized data are shown in the following section.

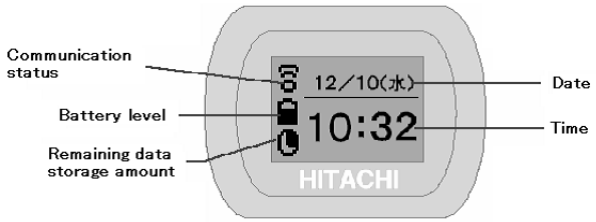



Fig. 1. Wearable wristwatch-type ZigBee sensor.

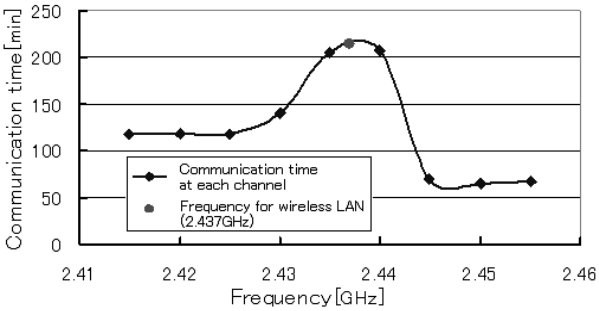



Fig. 2. Effect of co-channel interference on communication time.

## 3. FUNDAMENTAL PERFORMANCE OF ZIGBEE-BASED WRISTWATCH-TYPE SENSOR

Evaluations are carried out in order to clarify the fundamental performance of a wristwatch-type sensor complying with ZigBee. Firstly, an evaluation is carried out in order to clarify the effect of the co-channel interference on communication time. Next, the relationship between the communication distance and the throughput is evaluated.

Evaluated is the effect of co-channel interference of wireless LAN, which is based on IEEE802.11g, to the wristwatch-type sensor on communication time. Channel of wireless LAN is fixed to 6ch (2.437GHz) during the evaluation; besides the transmission power is 10dBm. On each channel of wristwatch-type sensor between 13ch (2.415GHz) and 21ch (2.455GHz), communication time until the end of transmission of all data stored on the data storage of wristwatch-sensor is measured. Figure 2 shows the communication time. The abscissa of Fig. 2 denotes the channel frequency[GHz] and the ordinate denotes the communication time[min]. From this figure, it can be seen that the communication time becomes longer when the channel frequency of wireless LAN is close to that of wristwatch-type sensor.

The effect of the distance between a wristwatch-type sensor and a receiver on throughput is evaluated under an indoor non-line-of-sight condition. Channel frequency is fixed during this evaluation. Throughput is measured at each position where the distance between a wristwatch-type sensor and a receiver is between 1m and 7m by 1m. Here, the throughput is calculated from the communication time until the end of transmission of all data stored on the data storage of wristwatch-sensor. Figure 3 shows the measurement result. The abscissa of Fig. 3 denotes the communication distance[m], and the ordinate denotes the throughput[kbps]. From this figure, although throughput is constant when the communication distance is between 1m and 6m; throughput measured at the distance of 7m is decreased because

TABLE II
SPECIFICATION OF WRISTWATCH-TYPE ZIGBEE SENSOR.

| Parameters | Contents | Remarks |
|---|---|---|
| Wireless communication system | IEEE802.15.4 | |
| Channel | 16ch(2405MHz∼2480MHz) | Communication distance ∼20m(Indoor) |
| Communication speed | 250kbps | Maximum speed |
| Transmission power | 0dBm | |
| Sampling interval for triaxial acceleration measurement | 20samples/s | |
| Sampling interval for temperature measurement | 1sample/s | |
| Data storage capacity | 14-day data (Maximum) | Capacity depends on usage environments. |
| Power supply | Internal battery（Rechargeable） | Charging time is about 2 hours. |
| Battery life | 14 days (Maximum) | Battery life depends on usage environments. |

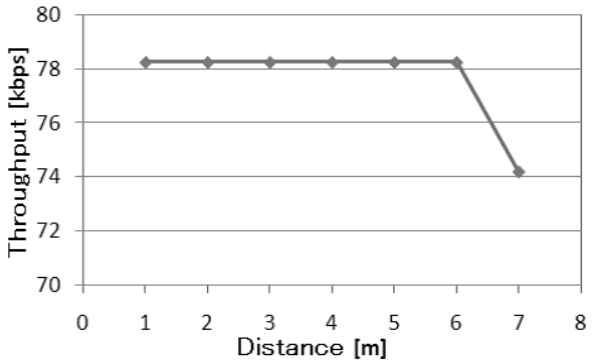



Fig. 3. Throughput vs. communication distance.

the communication link capacity is getting small.

Comparing the transmission power of ZigBee with that of wireless LAN, it is clear that the transmission power of ZigBee, or wristwatch-type sensor, is smaller than that of wireless LAN. Thus, wristwatch-type sensor which has small transmission power is easily affected by the co-channel interference. As a result, communication time becomes longer, and the communication link is easy to be disconnected when the distance between a wristwatch-type sensor and a receiver is over 7m.

## 4. NETWORK CONFIGURATION OF LIFE RECORDING AND MANAGEMENT SYSTEM

In Japan, the 2.5GHz frequency band has been allocated to two telecommunication carriers for wireless broadband systems at the end of 2007. Also, 10MHz frequency band called “local band” has been allocated to local carriers in order to improve public services or to eliminate digital divides. An experimental system connecting with the Internet using the “local band” has been deployed in Azumino city. Base stations complying with mobile WiMAX are deployed in 2009 and 2010, respectively. Problems in local cities are solved using the mobile WiMAX system. Concretely, current problems of informatization in local cities are shown as follows:

- P1: Expansion of the services demanded by citizens
- P2: Increase of the Internet utilization rate at home and edification of the effectiveness of Internet services
- P3: Elimination of digital divides
- P4: Improvement of information literacy

The following services will be provided using the mobile WiMAX network in order to solve the above-mentioned problems.

- S1: High speed Internet access service
- S2: Mobile VoIP (Voice over IP) telephonic service
- S3: Mobile VoD (Video on Demand) service
- S4: High quality image transmission service using live cameras
- S5: Life recording service
- S6: ICT disaster prevention system

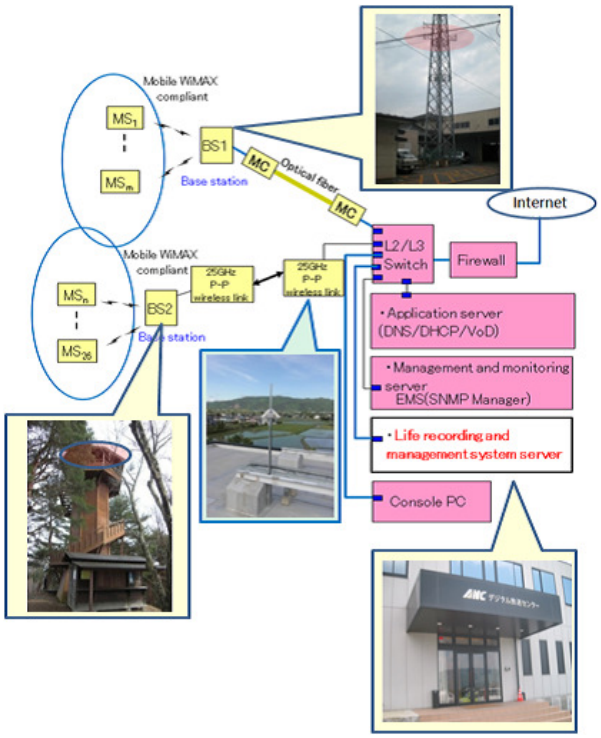



Fig. 4. Network architecture of the mobile WiMAX network in Azumino city.

Figure 4 shows the network architecture of the mobile WiMAX network deployed in Azumino city[5]. In this paper, the life recording and management system as one of the above services is introduced.

The life recording and management system consists of two networks as shown in Fig. 5. One is a ZigBee-based sensor network for collecting the data measured by a wristwatch-type sensor; the other one is a WiMAX-based infrastructure network for transferring the collected data to a server and for visualizing them.

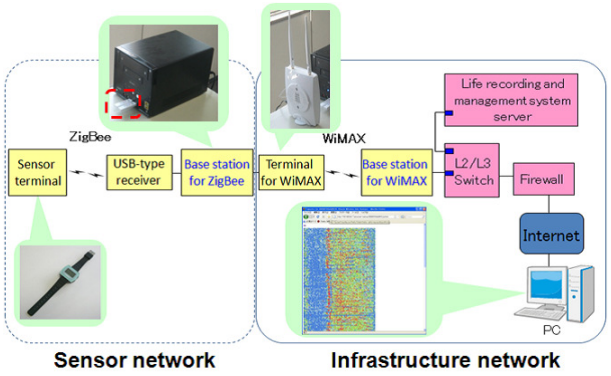



Fig. 5. Network architecture of the life recording and management system.

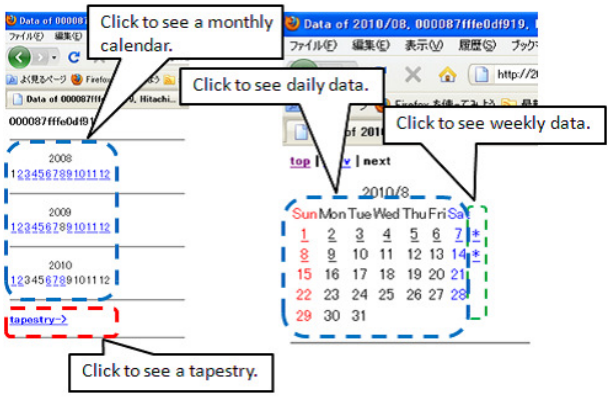



Fig. 6. Web pages for the life recording and management system. (Underlied figures indicate the existance of physical condition data)

### ZigBee-based sensor network

The network rounded by dotted line in Fig. 5 is a ZigBee-based sensor network. Communication between a wristwatch-type sensor and a USB-type receiver attached with a base station is performed by ZigBee. Amount of activity and temperature as physical condition data are transmitted to a base station. The base station for ZigBee connects by cable to a terminal for WiMAX, and the measured data are transmitted to a life recording and management system server via WiMAX network.

### WiMAX-based infrastructure network

The network rounded by solid line in Fig. 5 is a WiMAX-based infrastructure network. In this network, a base station complying with IEEE802.16e is employed. The frequency is 2587MHz which is an open frequency band for local communities and the band width is 10MHz. Transmission power of the base station is 33dBm per an antenna. The base station consists of two antennas for transmission and four antennas for reception and can perform antenna beam-forming and space diversity. The frame structure is OFDMA-TDD with 47 OFDM symbols per a frame. Frame interval is 5ms. Moreover, the symbol ratio between down link and up link is 29:18. A sub-channel allocation method is employed PUSC (Partial Usage of SubChannels). In a down link frame, 720 sub-carriers are split into 30 sub-channels. On the other hand, in an up link frame, 560 sub-carriers are split into 35 sub-channels.

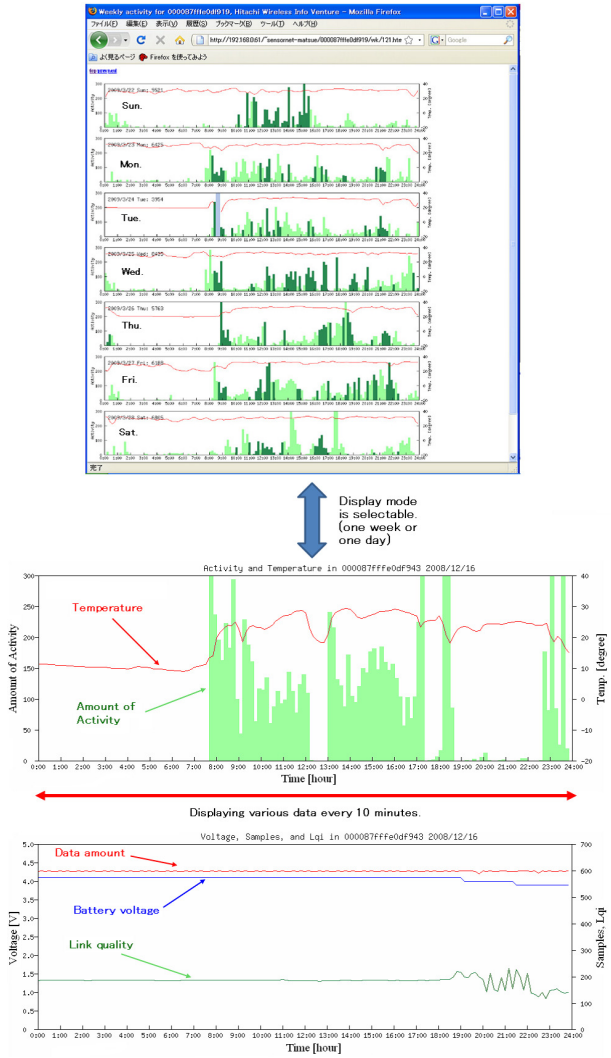



Fig. 7. An example of various physical condition data and sensor information (Upper: one week data, Bottom: daily data).

Firstly, a base station for WiMAX receives the physical condition data transmitted from a base station for ZigBee. Next, these data are stored on the life recording and management server; and life tapestry etc. are generated by the life information visualization software. Finally, users can confirm their own physical condition data by web browsers.

## 5. REPRESENTATIVE EXAMPLES OF VISUALIZED PHYSICAL CONDITION DATA

Measured data by a wristwatch-type sensor are analyzed by the life recording and management system server, and these data can be displayed by web browsers. Figure 6 shows web pages of the life recording and management system. In this page, users can check the daily data, weekly data and the life tapestory. Figure 7 shows an example of various physical condition data and sensor information of one week, upper, or one day, bottom. The bottom graph of Fig. 7 shows various physical condition data. The abscissa of this graph denotes the time, the left ordinate denotes the amount of activity which is displayed by a value between 0 and 300, and the right ordinate denotes the

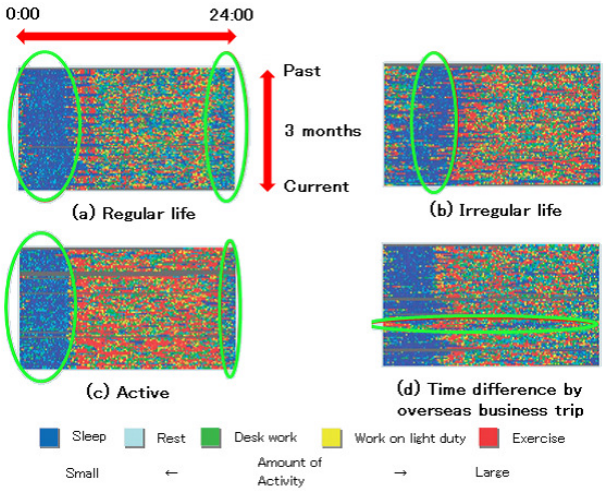



Fig. 8. Four examples of life tapestries.

temperature between -20 degrees and 40 degrees. The line plot indicates the temperature, and the bar graph indicates the amount of activity. From this graph, it can be seen that there is high correlation between the temperature and the amount of activity. Moreover, it can be seen that user removed the wristwatch-type sensor between 12:00 and 13:00 and between 18:30 and 22:30 because the temperature drops and there are little amount of activity.

The bottom graph of Fig. 7 also shows the sensor information. The abscissa of this graph denotes the time, the left ordinate denotes the voltage between 0 and 5.0, and the right ordinate denotes both the number of transmitted data samples between 0 and 700 and the Lqi (Link quality indicator) between 0 and 255.

Figure 8 shows four examples of life tapestries. Life tapestry indicates the amount of activity for 3 months, and it is painted by 5 colors. The abscissa denotes the time and the ordinate denotes the date for 3 months. From each life tapestry, the following things are clarified.

(a) Regular life: It can be seen that the user sleeps between 22:00 and 6:00 every day because the color of life tapestry is dark.

(b) Irregular life: It can be seen that the user stays awake until midnight because the light color exists at mid-night.

(c) Active: It can be seen that the user spends active life because the light color exists more than (a) between 6:00 and 23:00.

(d) Time difference by overseas trip: It can be seen that the user went to overseas trip during a week surrounded by circle because the user acts between 0:00 and 6:00, sleeps between 10:00 and 18:00.

If users wear the wristwatch-type sensor every day, they can correctly comprehend their own life rhythm[9][10][11].

## 6. RESULTS OF THE FIELD TRIAL TEST BY CITIZENS

In this section, results of the field trial test on the life recording and management system are shown. Total 151 citizens participated in the trial test. Firstly, the result of five-grade evaluation is shown in Table III. From Table III, the average evaluation score is about 3.5; consequently it is clear that citizens satisfy with the life recording and management system.

TABLE III
EVALUATION RESULTS OF 151 CITIZENS.

| | Evaluation | Rate (%) |
|---|---|---|
| 5 | Very good | 15 |
| 4 | Good | 35 |
| 3 | Normal | 38 |
| 2 | Poor | 12 |
| 1 | Very poor | — |

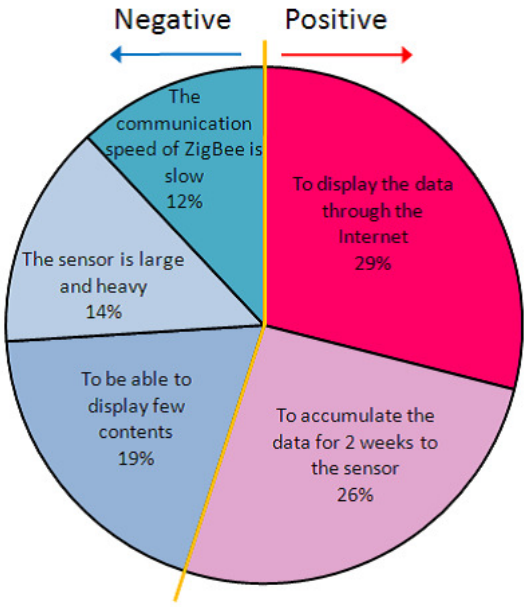



Fig. 9. Positive and negative comments from citizens.

Figure 9 shows comments from citizens which are largely classified into five types. From Fig. 9, about 55 % of the comments is positive one; on the other hand, about 45 % of them is negative one. The function of the system which was highly evaluated by citizens is to be able to confirm the history of physical condition data via the Internet. The secondly evaluated function is to be able to accumulate the measurement data for 14 days to the sensor terminal. On the other hand, some requests from citizens are to increase the contents sensed by the sensor terminal, to realize the more close-fitting sensor terminal and to accelerate the communication speed of sensor network.

## 7. CONCLUSION

In this paper, ZigBee-based life recording and management system is developed; and the overview, network structure and fundamental performance of the system are denoted. ZigBee-based life recording and management system automatically collects daily physical condition data, and it can visualize these data. Users can easily comprehend their own physical condition.

On the other hand, some problems for practical use have been clarified by evaluations. Firstly, the communication time becomes longer when the channel frequency of wireless LAN is close to that of wristwatch-type sensor. Secondly, wristwatch-type sensor which has small transmission power is easily

affected by the co-channel interference. As a result, communication time becomes longer. Finally, when the distance between a wristwatch-type sensor and a receiver is over 7m, the communication link is easy to be disconnected because the transmission power of a wristwatch-type sensor is small. Hence, users need to care about the distance from the appliance which emits the signal having same frequency with a wristwatch-type sensor. Or, users have to set the different channel frequency to each appliance.

In this paper, we realized the life recording and management system using the wearable wristwatch-type sensor as the sensing technology, both ZigBee and WiMAX as the wireless technology and the data visualization technology. The system fulfills three operational modes of wearable computing [12]. The first one is constancy. The system is always ready to interact with the user because the system can accumulate the physical condition data to the server using wireless technologies automatically within the range where the sensor can communicate with the base station. On the other hand, the system can also accumulate the sensor data for 2 weeks to the sensor even if the sensor cannot communicate with the base station. The second one is augmentation. In this system, the user is not required to control the sensor to sense and to transmit the physical condition data. Moreover, the user can wear the sensor as same feeling as when the user wears the wristwatch. So, the user gets relief from stress, and users can be doing something else at the same time as doing the computing. The third one is mediation. The system plays a role as mediator between a user and the outer world. In this system, the user's physical condition data, or internal data, is collected by the sensor automatically. And the system visualizes the user's physical condition data, and shares them through the Internet among the user and the people of the outer world, for example medical experts.

Currently, in Japan, PHR (Personal Health Recorder) project has been conducted by the Ministry of Internal Affairs and Communications; the Ministry of Health, Labor and Welfare; and the Ministry of Economy, Trade and Industry. This project aims to develop an infrastructure of electronic health record system. The data collected by the system includes the health information recorded at hospitals, pharmacies and homes. In this project, the life recording and management system shown in this paper is corresponding to a system which records the personal physical condition data at each home.

In the future, aging societies will increase the demand for systems like a life recording and management system. Thus, we will improve the system to be able to sense the various physical data through some wearable computer, and we will realize a system which can help the personal health care.

## 8. ACKNOWLEGEMENT

These studies were carried out in the SCOPE project presented by MIC of Japan.

## 9. REFERENCES

[1] IEEE Std 802.15.4-2006,"Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs)," 2006.

[2] ZigBee Alliance：
http://www.zigbee.org/

[3] Skyley Networks：
http://www.skyley.com/products/developmentkit.zigbeepro.html

[4] Renesas:
http://www2.renesas.com/micro/en/technology/zigbee/index.html

[5] H. Matsue et al., "Experimental Study of Networks and Services Using Mobile WiMAX System in Azumino City," RCS-2009-235, pp.163–168, Jan. 2010. (In Japanese)

[6] Hitachi Ltd. WirelessInfo Division
http://www.hitachi.co.jp/wirelessinfo/ (in Japanese)

[7] S. Yamashita, T. Shimura, K. Aiki, K. Ara, Y. Ogata, I. Shimokawa, T. Tanaka, H. Kuriyama, K. Shimada and K. Yano, "A 15 $\times$ 15 mm, 1 $\mu$A, reliable sensor-net module: enabling application-specific nodes," Proc. of the 5th international conference on IPSN '06, Apr. 2006.

[8] T. Tanaka, K. Aiki, H. Kuriyama and S. Yamashita, "A Low-Power Continuous Sensing Technique: Cool Sense," USN107(447), pp.61–64, Jan. 2008. (in Japanese)

[9] K. Suzuki, S. Yamashita, K. Aiki, H. Kuriyama and K. Shimada, "Key Technologies for Healthcare Sensor-Net Services - The Next Technology Frontier Towards Realworld IT -," Proc. of the 19th IEEE International Conference on MEMS 2006, pp.90 – 93, 2006.

[10] "Life Microscope: Future of Human, Sensor and IT that 20 Life Tapestries Tell," Hitachi hyoron, Vol. 89, No. 12, pp. 30–35. (in Japanese)

[11] T. Tanaka,S. Yamashita, K. Aiki, H. Kuriyama and K. Yano, "Life Microscope: Continuous daily-activity recording system with tiny wireless sensor," Proc. of the 5th INSS 2008, pp. 162–165, Jun. 2008.

[12] S. Mann, "Wearable Computing as means for Personal Empowerment", Proc. of the 1998 International Conference on Wearable Computing, ICWC-98, May 1998.

# Evaluation of studies at the Faculty

Tatjana KOVAC
Faculty of Commercial and Business Sciences
Celje, Slovenia

Marjana MERKAC SKOK
Faculty of Commercial and Business Sciences
Celje, Slovenia

**Abstract**

In its short career the Faculty of Commercial and Business Sciences (Celje, Slovenia) has created an important place in the Slovenian higher education space. Every year, more students enrol in the Faculty, which offers more courses. Today we have about 500 full time and 2500 part time students enrolled in three courses at all levels. Who are the Faculty's students? How successful are they in their studies and employment? What are their study experiences in the educational process at the Faculty? Does the study meet their expectations? These are questions which we try to analyze and answer in our paper.

The faculty held separate databases on the performance of students, successfulness of graduates and the satisfaction of students with their studies. Analysts preparing evaluation reports and using these resources have neither a single method nor the technology for extracting the relevant information. Reports are therefore non-transparent and do not say much about the quality of school work.

In the paper we will explain the applicability of the instrument (the expert model), as well as the methodology of its use. The instrument and the methodology may be useful for internal or external evaluation of studies at the Faculty.

**Key words**: faculty quality, evaluation of studies, expert model

## 1 INTRODUCTION

The Faculty collects numerous data and information about the study performance of students and graduates as well as about the satisfaction of students with their studies. These are separate databases. Analysts have problems compiling evaluation reports and analyses. They do not have the instruments or computer tools to prepare quality reports.

Reports and analyses are not related, as they contain data from separate databases. It cannot be deduced from them, for example, how study performance of students is related to successful and timely graduation or the satisfaction of students at the faculty. In order to be able to effectively monitor the quality of the organizational and educational work, it is necessary to establish how the the following three parameters influence the quality of work: performance of students, successfulness of graduates and the satisfaction of students with their studies.

We will therefore present an expert model designed in the computer programme Dexi [2], which enables to link data from different databases and present uniform information about the quality of studies at the faculty. It needs to be emphasized that the data reflects the opinion of students.

## 2 DATA SOURCES

The sample for the analysis comprised students who enrolled in the Faculty of Commercial and Business Sciences for the first time in the years 2007/2008, 2008/2009 and 2009/2010. The analysis included

206 full time and 1271 part time students. A profile of students at the Faculty was based on these data. The data were collected from the student database [5].

The sample for the analysis of graduates comprised students who graduated from the Factuly in the period between October 2007 and September 2009, i.e. in the study years 2007/08 and 2008/09. The analysis included 787 graduates. The data were collected from the graduate database.

The satisfaction with the Faculty's organization and management, the implementation of the education process, with teachers and other staff and with other activities of the Faculty is evaluated by students with the help of survey questionnaires. The satisfaction with the Faculty's organization and management is evaluated in respect of the adequacy of the Faculty's facilities, the work of the student affairs office, information on the Faculty's website, the organization of exams and the computer equipment in lecture rooms. The satisfaction with educational work is measured in respect of the quality of the student timetable, study contents and the quality of the implementation of lectures and practical classes (lecturers and other teachers). Some other factors also influence the satisfaction: integration of the faculty into the environment, motivation of student, coordination of work etc. The analysis took into account the survey results from 2007/2008 and 2008/2009 [6] and [7]. For the purpose of evaluating the Faculty's organization and management we analysed 220 questionnaires in the study year 2007/2008 and more than 3000 in the next study year.

### 2.1 Profile of students

We wanted to establish a profile of the students enrolled in our Faculty, which would take into account the following: gender, status (full time or part time), study year, the period between the completion of previous education and the enrolment in our Faculty, the school completed prior to enrolling in our Faculty and the results achieved in previous education.

In the analysed years the majority of full time students are female and their share is increasing by almost 10% each year. The ratio between female and male student is much more balanced among part-time students. A more balanced ratio in both study groups is desirable.

From year to year the enrolment in full time study is decreasing, which is the result of the decrease of enrolment in secondary schools (dwindling population). The enrolment in part time studies, however, is increasing (the result of life-long learning). The number of students enrolled in the first year is also on the increase, which indicates that less students transfer from other institutions of higher education. In the light of financial effects, this is desirable.

From the data for enrolled full time students we concluded that at least one year passes on average between the completion of previous education and the enrolment in our studies. The average grade of final examinations at previous education is improving, which means that each year better performing student enrol in our Faculty. On average, students enrol in part time studies two years after completing their previous education, but the actual data are much more dispersed than the data for full time students (some decide to study even after a 10-year break). The data on the grade of final examination at previous education are kept very inconsistently (21% of the sample includes this information), and can thus not be taken into account as a relevant indicator of the part time students profile.

Full time students enrol from 26 secondary and higher schools in the region, most of them from secondary schools of economics. Part time student enrol from a wide range of secondary schools in Slovenia (the records include 151 different secondary and higher schools).

### 2.2 Profile of graduates

The majority of graduates are female (71.4%), which is understandable, as 2/3 of the student population are female. On average, the study at our faculty lasts 2.5 years, which shows that many

students enrol directly to the second year (graduates of higher schools). The average final thesis grade is 9.31, while the grade point average is 8.18. In order to be able to provide quality remarks on the grade, a comparison with a similar faculty and comparable study programme would be needed. However, the final thesis grade is very high.

The analysis shows that approximately half of the graduates are employed, which is less than expected, bearing in mind the full time/part time students ratio at the faculty (14% full time and 86% part time students in the graduates sample).

Around 70% of graduates in the observed period graduated according to plan. 25% graduate in less than 2 years, 49 % in less than three and 5 % later.

Average final thesis grades are very high: 47.4% get the grade 10, 37.6% the grade 9, and 15% the grades 7 and 8. Grade point average is more evenly distributed: 46.6% score between 8 and 8.99, 36.7% between 7 and 7.99, and 16.7% get other grades. In comparison with final thesis grades, the grade point average is more realistic.

### 2.3 Data from student questionnaires

Educational work was assessed by around 2000 students in 2007/08 and by more than 3000 students in 2008/09. The average grade for teachers and study courses are almost identical in the observed period; 1.8 and of 2.3 respectively (on a scale from 1 to 4, where 1 is the highest and 4 the lowest grade). The grade for other teaching staff (teaching assistants, instructors, invited lecturers) varies more: 1.7 in 2007/08 and 1.6 in the next year.

The satisfaction with the Faculty's organization and management was assessed with the grade 1.9 by 220 student in 2007/08 and with the grade 2 by 357 student in 2008/09.

The other criteria, e.g. the integration of the faculty in the environment, motivation of full-time students, coordination of work, support of the management to the students, teacher training and adequacy of study literature, were assessed with the grade 1.8 in 2007/08 and 1.6 in 2008/09.

## 3 EXPERT MODEL

After processing the student questioners and analysing the data from the student database we were able to establish who our students and graduates are and how satisfied they are with the faculty. However, the interpretations of analysis results for individual study years are unrelated and non-transparent. Analysts lack the tool for comparing the results, examining the characteristic statistical differences in the evaluation of individual criteria and for simulating the desired conditions.

In what follows, we will present the expert model for evaluating the studies at our Faculty, which was designed with the computer programme Dexi (Decision Expert), we will make a sample evaluation for two observed periods, interpret the results and propose the use of the method.

Dex [3] is the shell of the expert system, intended for multi-parameter assessment. It provides effective assistance for decision-making regarding more complex problems [1]. DEXi is a derivate of the tool used in education. We evaluate individual study years according to criteria in the model, which are hierarchically structured; the grades for primary criteria are calculated by aggregation of grades for subordinate criteria according to decision-making rules. Besides an in-depth report, which can be compiled in Dex, the graphs and the so-called »what-if« analysis can help make simple simulations of the desired conditions.

In our case we built a model (prototype) for evaluating the studies at the Faculty, based on key criteria like: student profile, graduate profile, satisfaction of students (Figure 1). Criteria for establishing the profile of students (Par. 2.1), the profile of graduates (Par. 2.2) and the satisfaction based on the student questionnaires (Par. 2.3) were used as subordinate criteria. Average grades for criteria, calculated from previous analyses for the years 2007/08 and 2008/09, will be included in the model; grades for individual years present an individual variant assessed in the model.

The model prototype was tried out by a questionnaire among the faculty workers (model

validation). The sample included teachers, other Faculty staff and the management of the Faculty. The questionnaire was completed and handed in by 85.71% of the respondents. The discussion about the model validation can be extended and the model can be appropriately corrected. According to the collected opinions and suggestions, the Dex model may remain in the proposed form (Figure 1).

The grades for criteria in the years 2007/08 and 2008/09 were included in the model as variants. Two variants are enough to explain the usability of the method based on the interpretation of assessment results (Figure 2).

The evaluation of studies for both years is positive. However, there are certain differences which indicate whether or not the observed parameters are improving. Comparing the year 2008/09 with the previous one, the following differences may be established: the share of part time students increased (this is, according to employees, less favourable), the enrolment in the first year increased (which is desirable in the light of education work and the owners' interests), half of the graduates are employed and the other half are not (the previous year more graduates were employed), the integration of the faculty in the environment improved, as did the cooperation between teachers and students and between the management and students etc. If we were to examine grades for subordinate criteria, differences between both variants would be established. If the number of evaluated periods was higher (more variants) a trend could be observed – whether the evaluation of the study at the Faculty is improving or not; which differences demand special attention etc.

**Figure 1**
**Figure 2**

## 4 CONCLUDING REMARKS

This paper presents the method of evaluation of studies at the Faculty with the help of expert modelling. We used the computer programme Dexi, which is a simple programme in public domain, but any other comparable tool could also be used.

Expert model is a knowledge database, which can be used for establishing, monitoring and assuring quality at the Faculty. It collects data from dispersed databases at a single location, thus making the interpretation of results more transparent and enabling direct comparison of evaluations for different evaluated periods. The transparency of the evaluated variants (study years) makes it possible to react quickly and introduce improvements or changes.

The model may always be corrected by changing the relations and influences of individual criteria (correct decision-making rules). It would be reasonable to name a team of experts responsible for quality of studies, which would deal with such researches.

The presented method does not aim to automate the procedures and tasks related to the evaluation of quality at the Faculty. It present a sample support system for decision making, which collects data at a single location and helps experts to observe information which could previously be overlooked due to the dispersion of data. This is the significance and the value of the method.

## References

**[1]** Adelman, L. 1992. **Evaluating decision support and expert systems**. New York: John Wiley.
**[2]** Dexi**, Computer program for expert modelling**, available at http://kt.ijs.si/MarkoBohanec/dexi.html.
**[3]** Jereb, E., Bohanec, M., Rajkovič, V. 2003. **Dexi: računalniški program za večparametrsko odločanje**; uporabniški priročnik. Kranj: Moderna organizacija.
**[4]** Kovač, T. 2010. **Analysis of students and graduates at FKPV**. Celje: Faculty of Commercial and Business Sciences.
**[5]** **Educational records**. 2010. Celje: Faculty of Commercial and Business Sciences.
**[6]** Šuster Erjavec, H. 2008. **Students survays**. Celje: Faculty of Commercial and Business Sciences.
**[7]** Šuster Erjavec, H. 2009. **Students survays**. Celje: Faculty of Commercial and Business Sciences.

| The tree of criteria | Description of the criteria |
|---|---|
| **Evaluation the study on Faculty** | Evaluation the study on Faculty, the view of students |
| ├**Student profile** | Students who enroll in Faculty |
| │├**Full time students** | Full time students |
| ││├Percentage | Percentage of students |
| ││└**Entry properties** | Properties of students that enter the Faculty |
| │││├Continuing education | Number of years from the end of previous school |
| │││├Success of previous school | Success of previous school |
| │││└Enrollment in the class | Enrollment in the class; in the 1st or 2nd class |
| │└**Part time students** | Part time students |
| │├Percentage | Percentage of students |
| │└**Entry properties** | Properties of students that enter the Faculty |
| │├Continuing education | Number of years from the end of previous school |
| │├Success of previous school | Success of previous school |
| │└Enrollment in the class | Enrollment in the class; in the 1st or 2nd class |
| ├**Graduate Profile** | Graduate Profile of the Faculty |
| │├Percentage | Percentage of graduates in a year |
| │├**Success** | Success of diplomas and study |
| ││├Years of study | Years of study |
| ││├Assessment of study | Average assessment of study |
| ││└Assessment of diploma | Assessment of diploma |
| │└Employment | Student employment |
| └**Student satisfaction** | Student satisfaction; estimates from surveys |
| ├Organization and Management | Organization and Management; estimates from surveys |
| ├**Pedagogical process** | Evaluation of pedagogical work; estimates from surveys |
| │├Evaluation of subjects | Evaluation of study subjects |
| │├Evaluation of teachers | Evaluation of teachers; estimates from surveys |
| │└Evaluation of others | Evaluation of others in teaching process; estimates from surveys |
| └Other criteria | School and community collaboration, student motivation, student and teache |

**Figure 3: The model of evaluation studies at the Faculty, made in computer program Dexi.**

| Criteria | The year 2007/08 | The year 2008/09 |
|---|---|---|
| **Evaluation the study on Faculty** | good | good |
| ├**Student profile** | desirable | desirable |
| │├**Full time students** | desirable | desirable |
| ││├Percentage | ***planned*** | **to small** |
| ││└**Entry properties** | desirable | desirable |
| │││├Continuing education | ***immediately*** | after 1 year |
| │││├Success of previous school | good | very good |
| │││└Enrollment in the class | equally into 1st and 2nd class | ***mostly into 1st class*** |
| │└**Part time students** | desirable | desirable |
| │├Percentage | ***planned*** | to big |
| │└**Entry properties** | desirable | desirable |
| │├Continuing education | between 6 and 10 years | between 3 and 5 years |
| │├Success of previous school | good | good |
| │└Enrollment in the class | **mostly into 2nd class** | equally into 1st and 2nd class |
| ├**Graduate Profile** | ***very desirable*** | desirable |
| │├Percentage | ***planned*** | ***planned*** |
| │├**Success** | good | good |
| ││├Years of study | ***less than 3 years*** | ***less than 3 years*** |
| ││├Assessment of study | 8-8,99 | 8-8,99 |
| ││└Assessment of diploma | 9 | 9 |
| │└Employment | ***mostly employed*** | half employed-half unemployed |
| └**Student satisfaction** | good | good |
| ├Organization and Management | good | good |
| ├**Pedagogical process** | good | good |
| │├Evaluation of subjects | acceptable | good |
| │├Evaluation of teachers | good | good |
| │└Evaluation of others | good | ***exellent*** |
| └Other criteria | good | ***exellent*** |

**Figure 4: The result of evaluation of two consecutive academic years.**

# A Novel 3D Histogram Equalization Algorithm For Stacks Of Confocal Microscope Images

**Anna Romanowska-Pawliczek**
e-mail: anna@romanowska.pl
Department of Applied Computer Science and Modelling, Faculty of Metal Engineering and Industrial Computer Science, AGH University of Science and Technology, Kraków, Poland

**Mirosław Głowacki**
Department of Applied Computer Science and Modelling, Faculty of Metal Engineering and Industrial Computer Science, AGH University of Science and Technology, Kraków, Poland

**Piotr Pawliczek**
Department of Computer Science, Faculty of Electrical Engineering, Automatics, IT and Electronics, AGH University of Science and Technology, Kraków, Poland

**Zbigniew Sołtys**
Department of Neuroanatomy, Institute of Zoology, Jagiellonian University, Kraków, Poland

**ABSTRACT**

In this paper we present a method to solve a problem of brightness changes within a stack of images obtained by confocal microscope. A result of specimen scanning is a series of 2D images. Consecutive images show deeper sections of stained tissue. Within a stack of images a light attenuation increases with the depth of imaged focal planes. This well known physical phenomenon is a major obstacle that stands in the way of any CLSM image based analysis. Because light attenuation can be compensated by histogram modeling techniques, a straightforward solution is reached by histogram equalization of all images in a stack. The aim is to remodel shapes of histograms in such a way that the intensity histograms of the resulting images become uniform within all possible brightness values. Numerous histogram equalization methods have been developed in the last decade. In this paper we propose a novel 3D histogram equalization algorithm. It is tuned to balance contrast and brightness of images within one stack. In this approach we assume a linear correlation of brightness among adjacent stack slices. Unlike other known algorithms our approach concentrate more on preservation of information than improvement of visual aspects.

The presented algorithm was tested on real data as a module embedded in the system for automatic 3D reconstruction of brain glial cells. Because this novel method is a powerful and effective tool for contrast enhancement and achieving evenly balanced luminescence of confocal stack images it substantially contributes to the quality of fully dimensional cell reconstruction.

**Keywords:** computer vision, pattern recognition, image processing, 3D reconstruction, confocal microscopy

## 1. INTRODUCTION

Confocal Laser Scanning Microscopy (CLSM), due to its ability of non-invasive optical sectioning is considered to be one of the most important advances in optical microscopy. The sample analysis not only protects the specimen from damage but also creates images corresponding to a variety of optical sections within a certain volume of the sample [1]. Image information generated by CLSM does not contain signals originating from remote locations of the sample. This results in a highly precise plan of the acquired images.

Changes in cells' morphologies are observed in various pathological conditions such as neurodegenerative diseases or tumor transformations [2]. Therefore, methods that allow for more accurate quantitative analysis of such changes can be useful both in research into the mechanisms of development of such diseases, as well as in medical diagnostics. The attempt to explain the etiology of these phenomena requires analysis of each cell separately, yet with the context of surrounding tissues.

In medical examination, every day diagnoses and prognoses are based on visual evaluation of the analyzed specimen. Due to its increasing accessibility, CLSM is a device frequently used to gain insight into the characteristics of cell compartments or specific cell structures. Some of the main drawbacks encountered when using fluorescent CLSM are for example changed intensity information or certain restrictions in generating images. Improvement of the CLSM technique itself is therefore highly recommended.

In confocal microscopy one of the main problems encountered is noise [3]. As shown by Ortiz de Solorzano and coworkers [2] and Rodenacker and coworkers [4] if the objects that scatter and absorb light rays before those can reach the focal plane are present, image stacks obtained using confocal optical microscopy have lower intensities and tend to be more blurred.

Intensity inequality is even a bigger problem than noise when conducting a 3D analysis [5]. There are many factors that influence spatial intensity heterogeneity like fluorescent attenuation along the confocal axis [6], image acquisition factors [7, 8, 9, 10], photobleaching [11, 12], variations of illumination exposure rate, spatially uneven distribution of dye, spatial characteristics of illumination beams [13] and fluorochrome microenvironment [14]. It remains uncertain how to define the contribution of all the factors to the lower photon intensity effect for each individual image and study [15].

Losing image intensity is a generally recognized problem, however there is no single 3D correction method known that can be used and enable a truthful quantitative analysis for all studies [3, 16]. It remains a challenge to find such a compensation method that would help solve both spatial intensity heterogeneity and intensity noise problems and lead to the improvement of image acquisition steps.

A way to overcome light intensity attenuation in CLSM is to use histogram modeling techniques which change images in such a way that their histograms have a desired shape [17]. Histogram Equalization method (HE) is one of the easiest and most commonly used histogram modeling techniques known. In this method, gray levels of the equalized image are made as close as possible leading to a better contrast effect and a more detailed image of dark or bright regions. Pixel values are also modified by HE. The intensity histogram of the image becomes uniform, therefore the contrast of the output image is enhanced.

There are many various HE methods established [18, 19, 20, 21] and used to, e.g., correct spatial intensity heterogeneity in the lateral plane [17]. One of HE modifications called Adaptive Histogram Equalization (AHE) has applications in adjusting intensity variations locally by computing local histogram within spatially different focal planes [22]. Still, AHE remains highly sensitive to noise. An improved AHE method, the Contrast Limiting Adaptive Histogram Equalization (CLAHE), reduces noise amplification present in AHE.

In our method it is assumed that distribution of voxels for the ideal image is very similar, i.e. it has the same degree of brightness. This means that in an ideal case, all the 2D images within a stack should have a similar histogram. The greater part of each of the sections of the stack in the XY plane is the background with the same resolution regardless of the depth of the cross section, so this assumption is quite close to reality.

## 2. MATERIALS AND METHODS

In order to test the brightness balance algorithm presented in this paper, specimens from rat brain were used. They were prepared in the following way. The animals were anesthetized and perfused transcardially with 4% paraformaldehyde in phosphate-buffered saline. Brains were removed and post-fixed in the same fixative as before. Sections of 50–150 μm in thickness were cut with a vibratome (Leica VT1000S) and stained immunohistochemically with antibodies against glial fibrillary acidic protein (GFAP is a marker of brain astrocytes) and counterstained with DAPI to visualize cell nuclei. Details of the staining method were described in [23].

The confocal laser scanning microscope, Zeiss LSM 510 Meta, was used to examine the prepared specimens. The 3D images presented in this paper were obtained by a scanning series of 1024 x 1024 sections with 0.28 μm steps. During the scanning process, the Plan-Apochromat 63/1.4 Oil DIC objective was used and the pinhole diameter was set to 96 μm. The obtained stack consisted of 67 slices and was acquired as 12-bit depth images with resolution equal to 0.14 μm per pixel. Each scanning process was conducted two times, first one for GFAP fluorochrome excitation and second one for DAPI fluorochrome.

During image processing both stacks (DAPI and GFAP) were processed independently. In order to equalize the brightness of images within a stack approach proposed in this paper was tested. The method described below was implemented in the C++ language.

**Method of Linear Scaling with Histogram Matching (LSHM)**

The proposed approach focuses on the characteristics of the histograms. Method presented below is based on two assumptions:

- Histograms of neighbouring 2D images should be very similar, since these represent neighbouring cross sections of the stack.
- All the factors that affect the final brightness of the image are linear and operate on each pixel in the same way, regardless of the values of other pixels. The relationship between the values of pixels in the ideal image and the image obtained in the scanning can then be written as a linear function:

$$L_j(p) = a_j \cdot p + b_j \tag{1}$$

The values of the coefficients $a_j$ and $b_j$ are different for each image $I_j$. The variable $p$ is the value of a point on the resulting image which corresponds to a point of the value of $L_j(p)$ of the ideal image.

Based on these assumptions, we can conclude that differences in the histograms of two neighboring images $I_j$ and $I_{j+1}$ are caused by the fact that each of them is the result of the transformation of the actual (ideal) image by another linear function $L_j^{-1}$. This function is the reciprocal of $L_j$. Moreover, we assume that the ideal histograms should be almost identical. The described method relies on setting the transformation $L_j$ for each image $I_j$.

The number of pixels in image $I_j$ of a value $p$ will be denoted by the function $h_j(p)$ that is defined for integer arguments and takes integer values. It is a well-known definition of the histogram. Similarly, the cumulative histogram will be denoted by $H_j$. In this work it is assumed, that $H_j$ is defined as follows:

$$H_j : \mathbb{R} \to \mathbb{R}$$
$$H_j(p) = \sum_{i=0}^{\lfloor p \rfloor} h_j(i) + (p - \lfloor p \rfloor) \cdot h_j(\lfloor p \rfloor + 1) \tag{2}$$

The function $H_j$ is a non-decreasing function. This allows to define the inverse function $H_j^{-1}$:

$$H_j^{-1} : \mathbb{R} \to \mathbb{R}$$
$$H_j^{-1}(v) = min_p\left(H_j(p) = v\right) \tag{3}$$

Knowing the transformation of the image $L_w = a_w * p + b_w$, which is known or calculated in advance, we wish to find an analogous transformation $L_j = a_j * p + b_j$ for the adjacent image $I_j$. The aim is to find such values of $a_j$ and $b_j$ coefficients, that the image $I_j$ after transformation by the function $L_j$ has the most similar histogram to the histogram of the ideal image $I_w$ which has been transformed by the function $L_w$. The problem that must be addressed is the manner of determining the values of coefficients $a_j$ and $b_j$. The proposed solution is based on comparison of cumulative histograms $H_j$ and $H_w$. For the pixel values $p$ for which the histogram has a nonzero value, we try to match the corresponding pixel value from the second image. Then, for the designated in this way pairs of pixel values we seek for best fitted values $a_j$ and $b_j$. For the solution of this problem a global minimum of the following function was adopted:

$$E(a_j,b_j)=\sum_{p\in\mathbb{Z}}^{h_j(p)>0\wedge H_j(p)\le 0{,}9\cdot N}\left|L_j(p)-L_w(H_w^{-1}(H_j(p)))\right| + \sum_{p\in\mathbb{Z}}^{h_w(p)>0\wedge H_w(p)\le 0{,}9\cdot N}\left|L_w(p)-L_j(H_j^{-1}(H_w(p)))\right| = \sum_{p\in\mathbb{Z}}^{h_j(p)>0\wedge H_j(p)\le 0{,}9\cdot N}\left|a_j\cdot p+b_j-a_w\cdot H_w^{-1}(H_j(p))-b_w\right| + \sum_{p\in\mathbb{Z}}^{h_w(p)>0\wedge H_w(p)\le 0{,}9\cdot N}\left|a_w\cdot p+b_w-a_j\cdot H_j^{-1}(H_w(p))-b_j\right| \quad (4)$$

The variables $a_j$ and $b_j$ are the only searched values in the described equation. The values of coefficients $a_w$ and $b_w$ for the reference ideal image $I_w$ must already be known beforehand. In the above formula the only values $p$ that are taken into account are the ones of which cumulative histogram value has not exceed 90% of the pixels in the image. In this way the voxels with the highest values are omitted, i.e. those representing the objects. If for the analyzed image $I_j$ more than 90% of the pixels in that image have the same value, the value of this image coefficient values $a_j$ and $b_j$ are appropriate coefficients $a_w$ and $b_w$ of the reference image.

To compensate the brightness of images in the stack, it is necessary to select one of the images as a model image, to which the other images will be adjusted. For each processed stack a picture of the highest quality was chosen the model image. The values of its coefficients are taken $a_w = 1$ and $b_w = 0$. For the image lying directly under the proceeding one, the coefficients $a_j$ and $b_j$ are calculated. After coefficients assignation, this image becomes the reference image for another image below, for which the new coefficients $a_j$ and $b_j$ are then calculated. This operation is repeated for consecutive images until the end of the stack is reached. A similar procedure is applied to the images lying above the first reference image. The overall scheme of brightness equalization algorithm is as follows:

1. One of the slices from a stack is selected as a reference image and assumed as the primary model image $I_w$ ($a_w = 1$, $b_w = 0$). To select the most appropriate image, the algorithm described in [20] is used.
2. The image lying beneath the image $I_w$ is considered to be the image $I_j$.
3. The comparison of histograms of images $I_w$ and $I_j$, followed by designation of coefficients $a_j$ and $b_j$ – as a global minimum of the formula (4).
4. If the picture is not the last image of the stack, it is then considered to be the image $I_w$ with coefficients $a_w = a_j$ and $b_w = b_j$, and the algorithm returns to step 2.
5. The image selected in step 1 becomes again a primary model image $I_w$ ($a_w = 1$, $b_w = 0$).
6. The image lying above the image $I_w$ is considered to be the image $I_j$.
7. The comparison of histograms of images $I_w$ and $I_j$, followed by designation of coefficients $a_j$ and $b_j$ – as a global minimum of the formula (4).
8. If the picture is not the first image of the stack ($j > 0$), it is then considered to be the image $I_w$ with coefficients $a_w = a_j$ and $b_w = b_j$, and the algorithm returns to step 6.

As a result of the algorithm for each image $I_j$ we obtain the coefficients $a_j$ and $b_j$ of a stack through which the image should be transformed into. The operation of assigning new values of the pixels is performed on floating point numbers. The resulting stack of images with the actual pixels is transformed back to the input pixels' type.

The cost function presented as formula (4) may be rewritten as the following formula:

$$F(a,b)=\sum_{i=1}^{n}|c_{i0}+a\,c_{i1}+b| \qquad a\in[a_{min},a_{max}] \quad (5)$$

Variable $n$ is equal to the number of addends from formula (4). The variables denoted as $c_{i1}$ and $c_{i0}$ are constants. Their values are calculated from cumulative histograms $H_w$ and $H_j$ as it was described in formula (4). The searching solution corresponds to the global minimum of function $F(a,b)$. An additional assumption was made to simplify the problem – the value $a$ is assumed to take values from interval $[a_{min}, a_{max}]$. It corresponds to the brightness quotient between neighboring slices, so it may be safely constrained to values from interval $[0, 50]$. Several methods of minimization function presented at formula (5) are known, for example [24, 25, 26 or 27].

## 3. RESULTS

To verify the effectiveness of the method proposed in this paper, its results were compared with results of two other well-known algorithms. As the first of the reference methods the Histogram Equalization algorithm was chosen, denoted hereinafter as HE. Histogram equalization is intended to spread image gray levels over the entire scale and to allocate an equal number of pixels to each gray level. The effect of histogram equalization for human observers is usually a more balanced and better contrasted image. Applying this approach to solve the problem presented in the work, that is brightness alignment within the image stack, the HE algorithm is performed independently on each of the 2D images. The other method used for comparative purposes was the CLAHE method, known as Contrast Limited Adaptive Histogram Equalization, as described in [24]. Like HE algorithm, CLAHE is aimed to match the histogram of each cross section within a stack. However, the CLAHE algorithm preserves the local contrast of processed image well, so that generally provides visually better results than HE.

One of the properties, which demonstrates the high quality of the resulting image is contrast. In the present study to assess the quality of applied methodology the coefficient defined in [25], which is a measure of contrast per pixel, was applied. Its value is determined by the following formula:

$$C=\frac{\sum_{(x,y)\in I}\ \sum_{i,\ j\in[-1,0,1]}|I(x,y)-I(x+i,y+j)|}{N} \quad (6)$$

This coefficient is calculated for each slice independently. Its value corresponds to the average difference between adjacent pixels. For each images' pixel neighborhood, defined by a window of 3x3 pixels, is analyzed. We assume here that the value of pixels lying outside the image equals 0.

To verify the method proposed above raw stacks of images obtained from CLSM were processed. Values of coefficient $C$ (contrast per pixel) calculated for following slices of obtained stacks were collected on graphs presented at Figure 1. Images of cross sections from original and processed stacks were depicted at Figure 2. Images processed by HE algorithm are of incomparably high contrast when compared with images of the

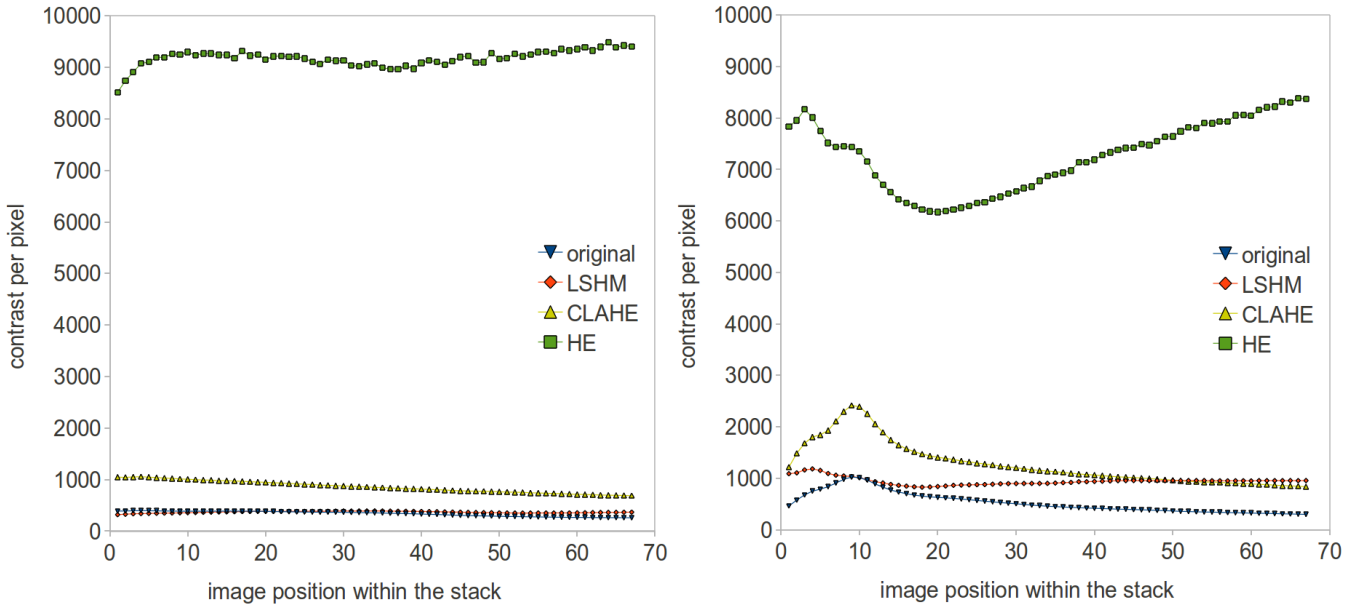



**Figure 1**. Contrast per pixel coefficient (formula (6)) calculated for DAPI (on the left) and GFAP (on the right) stacks.

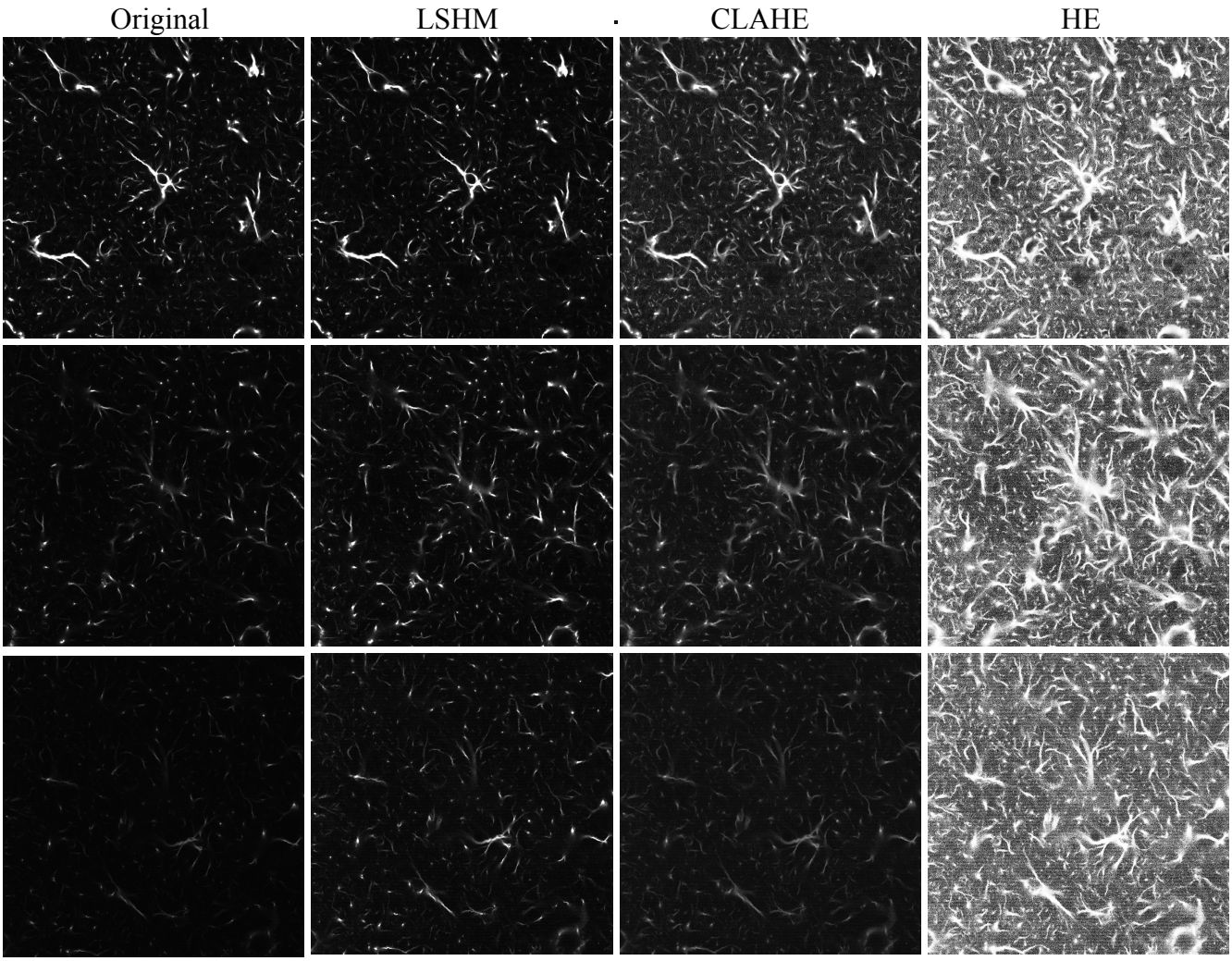



**Figure 2**. Images from GFAP staining stack (10th, 35th and 60th) before and after processing by LSHM, CLAHE and HE algorithm.

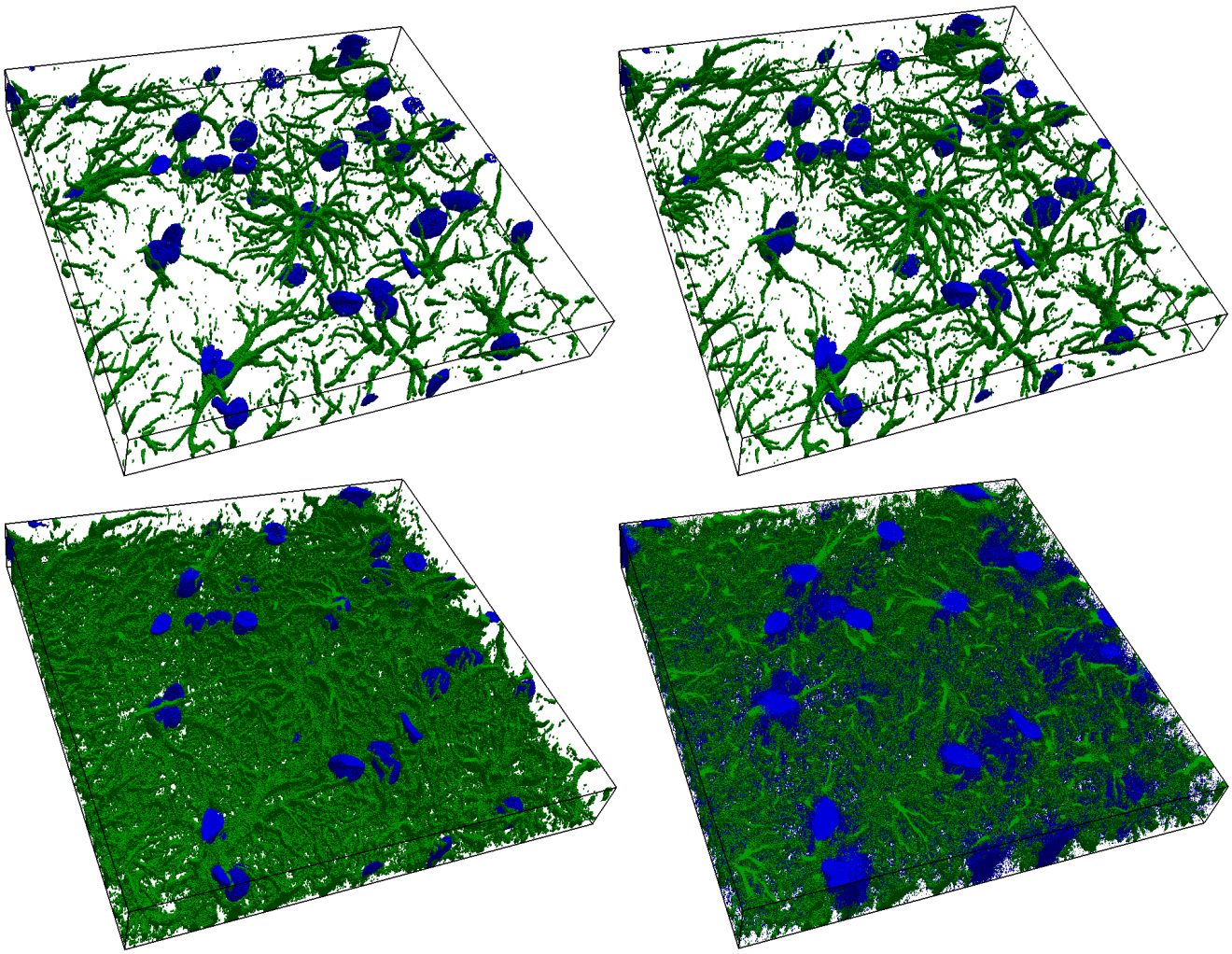


**Figure 3**. The results of Otsu binarization carried out on stacks: raw original data (top left), proposed algorithm of Linear Scaling with Histogram Matching (top right), CLAHE (bottom left), and HE (bottom right).

other stacks. It is caused by the fact that this method tends to equalize the histogram within each image. A side effect is a strong reinforcement of image noise of cross sections most of which is their background. This causes a big difference in the value of the coefficient C in comparison to the original stack (blue line) as well as to the stacks obtained by other methods. This is particularly clear for the stack of DAPI staining, which contains several nuclei distinctly separated against relatively uniform background. Application of HE algorithm for such a stack resulted in a strong differentiation of the background pixel values, representing the vast majority of the image, as reflected by the very high contrast ratio compared to the input stack.

Application of the CLAHE method to both tested stacks also caused explicit increase in the contrast, but to a much lesser degree. Unfortunately, the resulting shape of the graph is very similar to the one calculated for the original stack. This raises suspicion whether the application of this method does correct the difference in brightness of consecutive images within the stack or not.

Stacks processed with the use of the proposed method (red graph) have the lowest per-pixel contrast ratio. It is caused by the fact that our method places emphasis on preserving the shape of the input histogram. Both image stacks obtained by applying our algorithm are preferably aligned with contrast along the *Z* (depth) axis.

To verify the quality of processed stacks, they were binarized by a well known Otsu algorithm [28] and then visualized as a 3D image. In the Figure 3, the comparison between binarization results of unprocessed and processed stacks is presented. The green color marks voxels from GFAP stack, the blue color corresponds to voxels from DAPI stack. DAPI and GFAP stacks were processed independently and then merged.

As shown in Figure 3, binarization performed on the stack processed by our algorithm provides the best results. The quality of resulting 3D image is clearly better than the input image. The GFAP staining stack obtained by the CLAHE method also includes more details than the input stack. Nonetheless, due to large disproportions in the brightness of individual cross sections a huge fragment of the background is present in the result of binarization, causing a struggle with any further image processing. The DAPI staining stack was properly restored. The worst result was achieved by the use of the HE method. Because of strong deformation of histograms of consecutive images within a stack, the results of binarization incorporate a large amount of noise – that is in both the DAPI and GFAP staining stacks.

## 4. CONCLUSIONS

We have developed an intensity heterogeneity correction technique that adjusts the intensity heterogeneity of 3D image stacks and preserves fine structural details. The proposed method seems to be a promising alternative to initial processing

applied to CLSM image stacks. According to its exemplary use shown in the article, our method enabled significant quality enhancement of 3D images obtained from the processed stacks. It is true that its use may cause a deterioration of contrast to some sections of the processed stack. However, we expect this to do no harm in the context of three-dimensional processing, such as a full structure reconstruction. A big advantage of the proposed method is the simplicity of its application – it requires no ad hoc parameters settings nor any interaction with the user.

## ACKNOWLEDGEMENTS

We are grateful to Prof. E. Pyza, Dr. P. Weber and Mr. W. Krzeptowski from the Department of Cytology and Histology, Jagiellonian University, for providing the confocal microscope and support in our research.

This work has been supported by AGH University of Science and Technology under Grant No. 18.18.110.034.

## 5. REFERENCES

[1] T. Wilson, Confocal Microscopy: Basic Principles And Architectures., In: Confocal and Two-photon Microscopy: Foundations, Applications, and Advances, Diaspro, A. (ed.), Wiley-Liss, Inc., New York, 2001, pp. 19-38.

[2] C. Ortiz de Solorzano, E. Garcia Rodriguez, A. Jones, D. Pinkel, J.W. Gray, D. Sudar & S.J. Lockett, Segmentation of Confocal Microscope Images of Cell Nuclei In Thick Tissue Sections, Journal of Microscopy, Vol. 193, 1999, pp. 212-226.

[3] P. Paul, D. Kalamatianos, H. Duessmann & H. Huber, Automatic Quality Assessment for Fluorescence Microscopy Images, BIBE 2008, 8th IEEE International Conference on BioInformatics and BioEngineering, 2008. pp. 1-6

[4] K. Rodenacker, M. Aubele, P. Hutzler & P. S. Adiga, Groping for Quantitative Digital 3-D Image Analysis: an Approach to Quantitative Fluorescence In Situ Hybridization in Thick Tissue Sections of Prostate Carcinoma, Analytical and Cellular Pathology, 15 (1), 1997, pp.19-29.

[5] M. Jungke, W. von Seelen, G. Bielke, S. Meindl, et al., A System for The Diagnostic Use of Tissue Characterizing Parameters in NMR-Tomography. Proceedings of Information Processing in Medical Imaging (IPMI '87), 39, 1987, pp. 471-481.

[6] A. Semechko, R. Sudarsan, E. Bester, R. Dony & H. J. Eberl, Influence of Light Attenuation on Biofilm Parameters Evaluated from CLSM Image Data, Journal of Medical and Biological Engineering, 31(2), 2011, pp. 135-143.

[7] D.M. Benson, J. Bryan, A.L. Plant, A.M. Gotto Jr. & L.C. Smith, Digital Imaging Fluorescence Microscopy: Spatial Heterogeneity of Photobleaching Rate Constants in Individual Cells. Journal of Cell Biology, 100 (4), 1985, pp. 1309-1323

[8] M. Capek. J. Janacek and L. Kubinova, Methods for Compensation of the Light Attenuation with Depth of Images Captured by Confocal Microscope, Microscopy Research and Techniques., 69 (8), 2006, pp. 624-635.

[9] A.L. Franklin & W. G. Filion, A New Technique for Retarding Fading of Fluorescence: DPX-BME. Stain Technoogy, 60 (3), 1985, pp. 125-153.

[10] J. Rigaut & J. Vassy, High-Resolution 3D Images from Confocal Scanning Laser Microscopy: Quantitative Study and Mathematical Correction of The Effects from Bleaching and Fluorescence Attenuation in Depth. Analytical and Quantitative Cytology and Histology, 13 (4), 1991, pp. 223-232.

[11] V.E. Centonze & J.B. Pawley, Tutorial of Practical Confocal Microscopy and Use of the Confocal Test Specimen, In: Handbook Of Biological Confocal Microscopy, J.B. Pawley (ed)2005. New York, Springer Science+Business Media

[12] R.Y. Tsien, L.A. Ernst, A.S. Waggoner, Fluorophores for Confocal Microscopy: Photophysics and Photochemistry, In: Handbook of Biological Confocal Microscopy, J.B. Pawley (ed) 2005. New York, Springer Science+Business Media

[13] P. Van Oostveldt, F. Verhaegen & K. Messen, Heterogeneous Photobleaching in Confocal Microscopy Caused by Differences In Refractive Index and Excitation Mode. Cytometry, 32 (2), 1998, pp. 137-146.

[14] U. Tauer & O. Hils, Confocal Spectrophotometry. Scientific and Technical Information. (Special edition: Confocal Microscopy) CDR 4, 2000, pp. 15-27.

[15] K. Rodenacker, M. Hausner, M. Kuhn, S. Wuertz & S. Purkayastha, Depth Intensity Correction Of Biofilm Volume Data from Confocal Laser Scanning Microscopes, Image Analysis & Stereology, 20(Suppl. 1), 2001, pp. 556-560.

[16] J. Markham & J.A. Conchello, Artefacts in Restored Images Due to Intensity Loss in Three-Dimensional Fluorescence Microscopy, Journal of Microscopy, 204 (2), 2001, pp 93-98.

[17] R.C. Gonzales & R.E. Woods, Digital Image Processing, Upper Saddle River, NJ: Prentice-Hall, 2002

[18] J.H. Han, S. Yang & B.U. Lee, A Novel 3-D Histogram Equalization Method with Uniform 1-D Gray Scale Histogram, IEEE Transactions on Image Processing, Vol. 20 (2), 2011, 506-512.

[19] S.G. Stanciu & J. Friedmann, Compensating the Effects of Light Attenuation in Confocal Microscopy by Histogram Modelling Techniques, ICTON-MW'08; 2008.

[20] S.G. Stanciu, G.A. Stanciu & D. Coltuc, Automated Compensation of Light Attenuation in Confocal Microscopy by Exact Histogram Specification, Microscopy Research and Techniques., 73 (3), 2010, pp. 165-175.

[21] Q. Wang, L. Chen & D. Shen, Fast Histogram Equalization for Medical Image Enhancement, 30th Annual International IEEE EMBS Conference, 2008.

[22] S.M. Pizer, J.B. Zimmerman & E. Stabb, Adaptice Grey Level Assignment in CT Scan Display, Journal of the Computer Assisted Tomography, 8, 1984, pp. 300-305.

[23] P. Pawliczek, A. Romanowska-Pawliczek & Z. Soltys, Parallel Deconvolution of Large 3D Images Obtained by Confocal Laser Scanning Microscopy, Microscopy Research and Techniques, 73(3), 2010, pp. 187-194.

[24] I. Barrodale & F.D.K. Roberts, An Improved Algorithm for Discrete L1 Linear Approximation, SIAM Journal on Numerical Analysis 10 (5), 1973, pp. 839–848.

[25] E.J. Schlossmacher, An Iterative Technique for Absolute Deviations Curve Fitting, Journal of the American Statistical Association. 68 (344), 1997, pp. 857–859.

[26] G.O. Wesolowsky, A New Descent Algorithm for The Least Absolute Value Regression Problem, Communications in Statistics, Simulation and Computation B10 (5), 1981, pp. 479–491.

[27] Y. Li & G.R. Arce, A Maximum Likelihood Approach to Least Absolute Deviation Regression, EURASIP Journal on Applied Signal Processing 2004 (12), 2004, pp. 1762–1769.

[28] Nobuyuki Otsu, A Threshold Selection Method from Gray-Level Histograms, IEEE Transactions on Systems, Man & Cybernetics 9 (1), 1979, pp. 62–66.

# Image Recognition, Identification and Classification Algorithms in Cast Alloys Microstructure Analysis

**Anna Romanowska-Pawliczek**
e-mail: anna@romanowska.pl
Department of Applied Computer Science and Modelling, Faculty of Metal Engineering and Industrial Computer Science, AGH University of Science and Technology, Kraków, Poland

**Aleksander Siwek**
Department of Applied Computer Science and Modelling, Faculty of Metal Engineering and Industrial Computer Science, AGH University of Science and Technology, Kraków, Poland

**Mirosław Głowacki**
Department of Applied Computer Science and Modelling, Faculty of Metal Engineering and Industrial Computer Science, AGH University of Science and Technology, Kraków, Poland

**Małgorzata Warmuzek**
Foundry Research Institute, Kraków, Poland

**ABSTRACT**

Automatic identification and classification of objects being the results of image recognition algorithms became more and more popular in many aspects of human activity. On the other hand, manual stereological methods and conventional image analyzer are more often than not difficult and time-consuming tool to obtain the informative data especially for complicated microstructures. To solve these problems, a computer assisted quantitative metallographic analysis was explored. The input data for the proposed analysis was a set of digital 2D images of metal microstructures of the technical aluminum cast Al-Si alloys. Images were obtained by high quality cameras embedded in optical microscopes. The objects of interest were the precipitates of intermetallic phases of various morphological shapes.

Traditionally, descriptions of microstructures have been based on measurements of topological relationships between the three-dimensonal space and two-dimensional microsections, such as grain size, the average volume of particles, volume fraction, size of particles in unit volume, etc. We consider these features to be insufficient for the process of classification which permits differentiation. Therefore, the computational methods of pattern recognition have been applied to both the statistical particle shape analysis and topological characterization of dendritic structures. Several examples of designed and implemented algorithms, including the measurements of compactness, scale and rotation invariant moments, fractal dimension, convex hull, lacunarity and many other parameters are presented. The key to this quantitative analysis is the manner of interpretation of aluminum alloys' planar microsections. It provides practical techniques for extracting quantitative information from measurements. It is these features that determine the mechanical properties, and any advanced understanding of microstructure-property relations requires their quantitative description.

The presented approach is aimed at designing a system for identification and classification of microstructures occurring in multiphase cast alloys. Image data representing diverse samples was taken into investigation. Within each sample alloys' features were determined based on a cast modeling process. Due to the fact that the presence of specific microstructures determines mechanical properties of cast alloys, an automated image based classification system may be an invaluable tool for developers of modern casting technology.

**Keywords:** computer vision, pattern recognition, image processing, identification of metal phases, quantitative metallography

## 1. INTRODUCTION

Polyphase metal alloys are still the most common structural materials in the production of many goods. Increasing demands regarding on one side, the quality of the products and on the other saving costs and environmental friendly technology are opening the wide fields for advanced methods of material investigations. The starting points of prospective material modification are always very well-known relationships {C,T} ↔ {UP} (where C- chemical composition, T- technology, UP – utilizable properties) [1-6]. However, the another relationships {M} ↔ {UP} (where: M- microstructure) represents a more close and direct interaction which can be implemented into physical or statistical material models [7]. The term 'material microstructure' in material science means a 3D construction composed of the particular elements differing in physical, chemical and morphological properties. Light microscopy investigations relate to the 2D representatives of the microstructure constituents, revealed on the metallographic plane cross sections with special preparation procedures. The known stereology relationships allow direct matching of the 2D quantitative global parameters for some microstructure models to their 3D equivalents [8-10]. However, the general description rules for the local features of the material constituents, important from the point of view of its model behavior have not been until yet established [11-13]. Especially, in the case of concave dendritic particles an anticipated 2D ↔ 3D morphology relationship can be univocal and even contradictory (Fig.1).

The quantitative description of the local microstructure features as shapes of particular elements is one of the most important and difficult problems in microscope image analysis. A

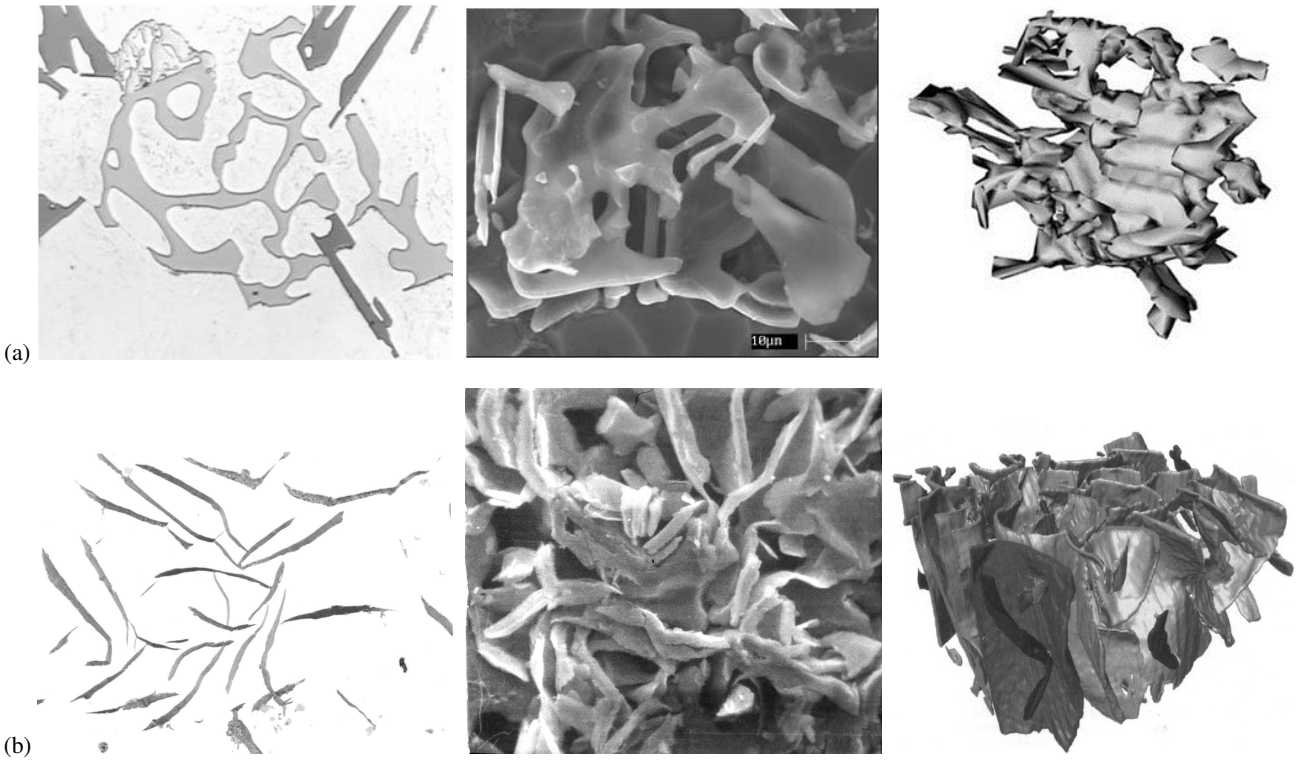



**Figure 1**. Results of the 3D reconstruction of the microstructure constituents shape: (a) sequence of the localized cross sections, intermetallic phase in the Al-Si alloy [12], (b) FIB in SEM, flake graphite in the cast iron [13] (1- LM, plane cross section (FRI microstructure archives), 2- SEM, deep etched cross section (FRI microstructure archives), 3- 3D reconstruction)

computer assisted microscope image analysis has been explored in order to solve this problem [8, 9, 14, 15].

In the multi component eutectic microregions during alloy crystallization the phase constituents solidify into very morphologically complicated forms. Commonly, concave solid figures of a well-developed surface are found, sometimes similar to either dendrite or fractal forms. On the basis of the metallography knowledge, the characteristic morphology forms are correlated to a particular phase constituent identified univocally by its crystal structure [16, 17]. Local crystallography phase identification is more time-consuming than microscopic observation, thus the quantitative procedure for classifying and discriminating observed morphology forms will be a very useful examination tool.

| shape | phase | crystal | element |
|---|---|---|---|
| needle | β-AlFeSi<br>$Al_3Fe$<br>$Al_7Cu_2Fe$ | monoclinic<br>tetragonal | Al, Si, Fe,Cu |
| branch | $Al_3Ni$<br>$Al_9FeNi$<br>$Al_6Cu_3Ni$ | orthorhombic<br>cubic | Al, Ni, Fe, Cu |
| Chinese script | α-AlFeSi<br>α-AlMnSi<br>$Mg_2Si$ | hexagonal<br>cubic | Al, Si, Fe,<br>Mn, Mg |

**Table 1**. Morphology forms established as shape standards for chosen intermetallic phases [16, 17]

Three morphology groups of the intermetallic phase precipitates occurring in the eutectic solidifying in the Al alloys have been chosen for the examinations. They have been named by with using wide analogies as needles, branches, Chinese script. Each of them is specific for a particular intermetallic phase group (Tab. 1), therefore the establishment of the quantitative coefficient allowing their separation on the microstructure images will be considered an important progress in the solution of the problem formulated above.

## 2. MATERIALS AND METHODS

**Material preparation**

The analyzed images represent microstructures of the technical aluminum cast Al-Si alloys. The microstructure examinations have been carried out on metallographic microsections, polished with 0,25mm diamond suspensions and etched with 10% NaOH.

Metallographic specimens of alloys' samples are subjected to digestion, which reveals the overall picture of the structure and enables identification of individual structural components. Reagents reveal the structure of etched phases and grain boundaries. Having a selective reagents' stain or dissolving certain components of the structure allows their identification.

The microstructure observation has been carried out by means of the light metallographic microscope either Neophot 32 or Axio OZm1. Microstructure pictures have been recorded in the digital form as jpg files.

**Image acquisition**

Materials of examinations were series of the laboratory cast multicomponent Al alloys. Microstructure images have been revealed with the standard metallographic procedures on the random plane cross sections. The pictures have been recorded in RGB standard with 36bit color depths by means of the light metallography microscope AxioObserver oZm combined with high resolution Axiocam ICc3 camera of CCd basic resolution 2080x1540.

The quality of images depends on the use of light source, filters, way of lighting the cross sections and choice of lenses mounded in the microscope. Pictures of microstructures taken at various magnifications include structural components (objects) which differ from the matrix by luminance and chroma. The analysis of such structures can be additionally impeded by uneven illumination and low contrast. It requires balancing of global and local histograms of images. Another problem is the lack of continuity of grain boundaries and phase brightness due to the heterogeneous matrix and the presence of noise in the image. Finding the location of boundaries between the phases in the image of the microstructure requires the use of various methods of detection depending on the nature of the structure and the lighting conditions.

**Image analysis**

Computer analysis of digital images requires finding a solution to many problems. There is no methodology that allows to approach each issue in the same way. For the analysis of microstructures of aluminum alloys, it is necessary to carry out a series of image operations, which result in the calculation of parameters describing the structural components.

On the base of the previous works [7, 16, 17], the morphology class of the observed microstructure constituents has been arbitrarily established. The real material constituents, represented on the microphotographs, i.e. specific image fields recognized as particular intermetallic phase precipitate representatives, have been attributed to the particular morphology class according to their a priori visual pattern recognition. The geometry of each morphology class of the objects ought to be recognized and univocally described by means of either one or another chosen group of the quantitative coefficients.

The analyzed objects are characterized by different structural shapes. It is important to describe shapes using parameters (features) whose values do not depend on the microscope magnification. From a mathematical point of view, it is insufficient to describe the shape only by a single feature. From the need to reduce the accuracy of shape description, it is required to propose and design certain shape specifications. These must be parameterless, easy to interpret and give values reflecting differences in the shape of a specific type of structure. For such purposes the use of so-called moments of inertia, topological parameters and objects boundaries analyzes seems promising. For each image these parameters can be calculated by finding coherent components and skeletons of the images' objects.

Description of shapes with a high degree of complexity causes some of the parameters to be similar for various classes of objects. In such cases, it is necessary to analyze a large sample of images so that the classification is carried out by using the average expected values.

**Moments of inertia**

In the process of image analysis the moments of inertia are very widely used [18]. They describe the image content or its distribution relative to the coordinate system. Moments reflect the change of global and local geometry of the structure. By analogy to the mechanics, image properties are characterized by moments. Assuming a two-dimensional image as a continuous function of the density distribution *f(x, y)*, the moment of order *(p+q)* for the entire image area $\Omega$ is defined as:

$$m_{pq} = \iint_{\Omega} x^p y^q f(x, y)\,dxdy \tag{1}$$

for *p, q = 0,1,2, ...* . As can be seen of Eq. (1) in the case of image processing, the moment is a special feature of the weighted average intensity of pixels. For binary images Eq. (1) is converted into a discrete form. For simplicity it is assumed that the image area is divided into squares of size 1x1, where the value of the density function is constant. To an image recorded in grayscale of brightness of pixels *f(x,y)*, the moment of *ij* is calculated as:

$$M_{ij} = \sum_x \sum_y x^i y^j f(x, y) \tag{2}$$

where *x* and *y* are the coordinates of successive pixels in the image. For images of alloys microstructures analyzed in the presented study, the moments were calculated according to Eq. (2). Because the microstructures were digested with various reagents and the light settings in the microscope were not constant the binarization was performed. Specific moments of binary image are: surface area, center of mass, orientation. For example, the image property described by moments expressed as:

- $M_{00}$ – the object's surface
- $M_{10}$ / $M_{00}$ – coordinate $x_c$ of gravity center of an object
- $M_{01}$ / $M_{00}$ – coordinate $y_c$ of gravity center of an object

Central moment *pq* for the image stored in the grayscale is defined as:

$$\mu_{pq} = \sum_x \sum_y (x - x_c)^p (y - y_c)^q f(x, y) \tag{3}$$

where *f(x, y)=1* for pixels representing objects of a binary image.

Central moments of the third row are calculated by formulas:

$$\begin{aligned}
&\mu_{00} = M_{00}\\
&\mu_{01} = \mu_{10} = 0\\
&\mu_{11} = M_{11} - x_c M_{01} = M_{11} - y_c M_{10}\\
&\mu_{20} = M_{20} - x_c M_{10}\\
&\mu_{02} = M_{02} - y_c M_{10}\\
&\mu_{21} = M_{21} - 2x_c M_{11} - y_c M_{20} + 2x_c^2 M_{01}\\
&\mu_{12} = M_{12} - 2y_c M_{11} - x_c M_{02} + 2y_c^2 M_{10}\\
&\mu_{30} = M_{30} - 3x_c M_{20} + 2x_c^2 M_{10}\\
&\mu_{03} = M_{03} - 3y_c M_{02} + 2y_c^2 M_{01}
\end{aligned} \tag{4}$$

Based on the above mentioned central moments Hu [19] proposed the two-dimensional image set of invariant moments. Seven independent moments can be used to identify and classify objects regardless of their size, position and rotation. For this purpose, normalized central moments $\eta_{ij}$ are defined:

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{00}^{\left(1 + \frac{i+j}{2}\right)}} \tag{5}$$

Moments regardless of scale, position and rotation $\Phi_i$ are recorded as a combination of moments $\eta_{ij}$:

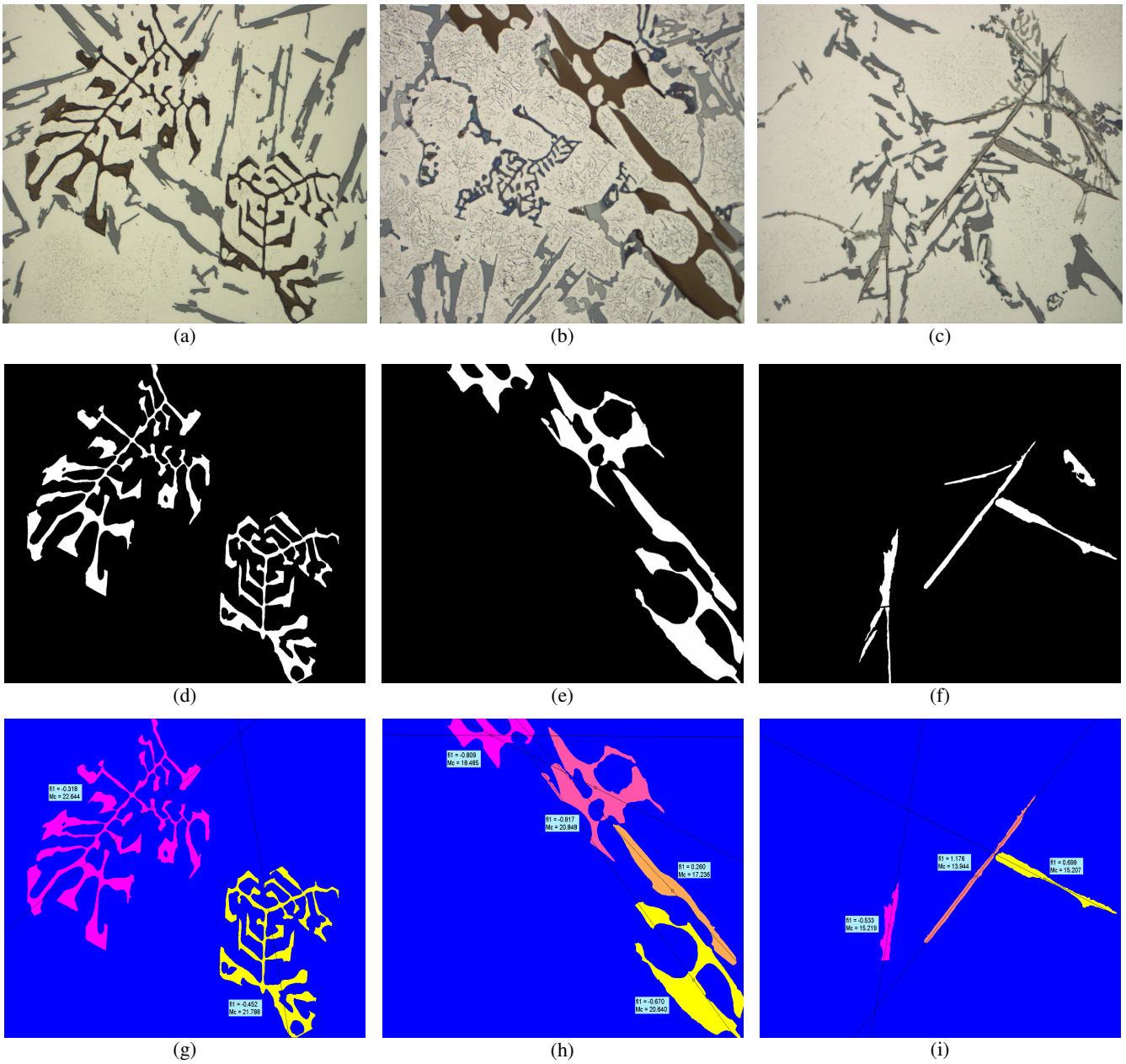



**Figure 2**. Overview of microscopic images of aluminum alloys at a magnification of 500x and images resulting from their processing. The first row comprises of pictures of microstructures such as: (a) Chinese script, (b) branches and (c) needles. The binarization process results in images presented in the second row (d, e, f). The last line contains images generated by segmentation process (g, h, i). Color indicates a blue background while further objects are marked with contrasting colors. The resulting images include information on: the objects gravity center position, the main axes, values of the first scale independent moment $\Phi_1$ and moments of inertia with respect to the main axis $M_C$.

$$
\begin{aligned}
\Phi_1 &= \mu_{20} + \mu_{02} \\
\Phi_2 &= (\mu_{20} + \mu_{02})^2 + (2\mu_{11})^2 \\
\Phi_3 &= (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \\
\Phi_4 &= (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \\
\Phi_5 &= (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})((\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2) \\
&\quad + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})(3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2) \\
\Phi_6 &= (\mu_{20} - \mu_{02})((\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2) \\
&\quad + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}) \\
\Phi_7 &= (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})((\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2) \\
&\quad - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})(3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2)
\end{aligned}
\tag{6}
$$

## 3. RESULTS

Acquired images of specimens' sections were binaryzed with the use of Otsu algorithm [20]. From obtained images 21 largest structures were chosen and classified by expert. Each object was assigned to exactly one of the following classes: needles, branches or Chinese script.

The aim of this study was to propose quantitative indicators describing the phases of the microstructure that will allow you to assign it to one of the above classes. Due to the shape of the phases occurring in the various structures, the best parameters should be such, that their value of which strongly depends on the shape. For each analyzed object moments $\Phi_1$-$\Phi_6$ were calculated. However, no combination of these parameters

allowed classification of analyzed structural components. Therefore, it became necessary to introduce a parameter sensitive to the morphological differences of metal phases. In this paper, a new morphological parameter moment of inertia of objects according to their main axis was introduced. This value was calculated by formula:

$$M_C = \sum_x \sum_y d_c^2 \square x, y \square f \square x, y \square \quad (7)$$

where $d_c(x,y)$ is the distance from the object pixel with coordinates $(x,y)$ from its main axis, and $f(x, y)=1$ for object pixels in the image binary.

The examples of structures from different classes and results of their processing are presented on Fig. 2. The best results in terms of their use in identifying the components of phase microstructures give: first moment $\Phi_1$ and moment of inertia with respect to the main axis of the objects $M_C$.

Fig. 3 is a graphic summary of the results of measurements of parameters $\Phi_1$ and $M_C$ for all 21 processed structures. Each marker was labeled by color that corresponds to one of the class assigned by expert: Chinese script (blue), branch (violet) or needle (red). Selection of the largest structural elements for each type of microstructures allowed to find a range of values for different parameters and minimize the measurement error value. The outlined areas, typical for ranges of values of particular structures parameters are depicted in Fig. 3.

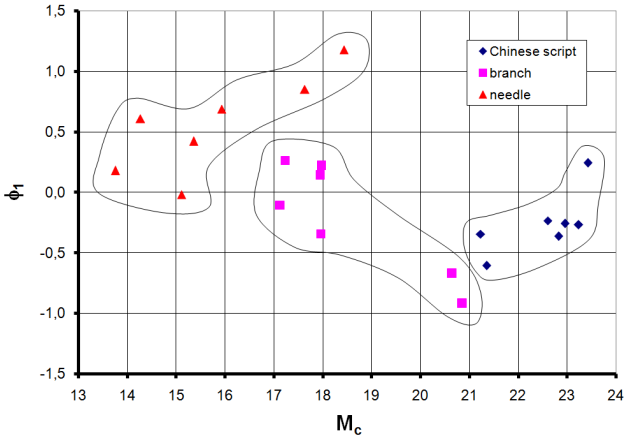



**Figure 3**. Summary of measurement results of parameters for the largest structures of the phases such as: Chinese writing (blue), branches (violet) and needles (red).

## 4. CONCLUSIONS

The aim of this study was to propose such quantitative indicators that describe the phase microstructure and allow to assign it to one of the three types of structures.

The results of the carried out quantitative microscope image analysis have revealed that more complicated morphology forms, present in the microstructure images of the cast Al alloys cannot be univocaly described with only one geometry shape factor used for tested images sets. Nevertheless, the presented experiment has shown the possibility of particular morphology class discrimination according to complex coefficients combining particular geometry shape factor with one of the binary image momentum. This result exposes to view the new field of quantitative microstructure description as a very important stage of the material model simulation and its technical application.

The positive verification of the assumed attribution of the morphology classes parameters to the particular microstructure constituents provides a new tool of computer aided microscope image interpretation.

Strong influence of both image quality (i.e. either metallography cross section preparation or acquisition conditions) and microscope magnification during present examinations suggests the necessity of this procedure stage standardization.

## ACKNOWLEDGEMENTS

This work has been supported by AGH University of Science and Technology under Grant No. 18.18.110.034. It has also been carried out with a financial support of the Polish Ministry of Science and Higher Education under grant No. NN507 378735.

## 5. REFERENCES

[1] M. Hafiz &T.Kobayashi, Metallurgical factors affecting impact toughness of eutectic Al-Si alloy, Z. Metallk. V89, 1998, pp.445-449.

[2] P.N Crepeau & S.D Antolovich, Structure property relationships in aluminium alloy 339-T5: tensile behaviour at room and elevated temperature, AFS Transactions, 1990, pp. 813-822.

[3] D. Lassance, Modelling of damage mechanisms in AlMgSi alloys, Pres. Universitaire de l`UCL, Belgique, book, 2003.

[4] Yun Sun Han, Effect of intermetallic phases on the mechanical properties of cast A356 alloy wheels, The 1st International Forum on Strategic Technology, 12-20 oct 2006, pp.338-341.

[5] M. Nylen, U. Gustavsson, B. Hitchinson & A. Ortinas, Mechanistic studies in clad aluminum alloys, Mat. Science Forum, Vol. 217-222, 1996, pp.1703-1708.

[6] R.G. Hamerton, H. Cama & M.W. Meredith, Development of the coarse intermetallic particle population in wrought aluminium alloys during ingot castings and thermomechanical processing, Mat. Science Forum, V331-337, 2000, pp.143-154.

[7] M. Warmuzek, Microstructure aspect of the usable properties formation of Al cast alloys (Mikrostrukturalny aspekt kształtowania użytkowych własności odlewniczych stopów Al), Odlewnictwo –Nauka i Praktyka, No.5-6, 2006, pp. 27-35.

[8] J.Ohser, 3D Images of Material Structures – Processing and Analysis, Wiley VCH, 2009, pp. 79-148.

[9] L.Wojnar, Image Analysis Applications in Materials Engineering, N. York, CRC Press, 1999.

[10] J.C. Jacquet & W. Hotz, Quantitative description of the microstructure of aluminium foundry alloys, Cast Metals, V4, 1992, pp. 200-225.

[11] H. Singh, A.M. Gokhale, Y. Mao, E.Tewari & A.K. Sachdev, Reconstruction and quantitative characterization of multiphase, multiscale three dimensional microstructure of a cast Al-Si base alloy, Met.and Mat.Trans. B, V40B, 2009, pp. 859-870.

[12] A. Velichko, Quantitative 2D Characterization of Graphite Morphologies in Cast Iron using FIB Microstructure Tomography, Dissertation, Saarbrucken, 2008, pp. 52-55, 95-111.

[13] C.M. Dinnis, J.A. Taylor & .A.K. Dahle, Interaction between iron, manganese and the Al-Si eutectic in Al-Si alloys, Met.and Mat Trans.A, V37A, 2006, pp. 3283-3291.

[14] C. Sigli, H. Vichery & B. Grange, Computer assisted metallurgy for non heat treated aluminum alloys, Mat. Science Forum, V212-217, 1996, pp. 391-396.

[15] W.R. Cribb, Quantitative metallography of poly-phase microstructures, Scripta Met. V12, 1978. pp.893-898.
[16] M. Warmuzek, Metallographic Techniques for Aluminium and Aluminium Alloys, in: Metallography, V9, ASM International, USA, 2004, pp. 711-752.
[17] J.E. Hatch, Aluminum – properties and physical metallurgy, ASM International, Metals Park, Ohio 1993, pp. 30-31.
[18] R. Mukundan & K.R. Ramakrishnan, Moment functions in image analysis: theory and applications, e-book, 1998.
[19] M.K. Hu, Visual Pattern Recognition by Moment Invariants, IRE Trans. Info. Theory, Vol. IT-8, 1962, pp. 179-187.
[20] Nobuyuki Otsu, A Threshold Selection Method from Gray-Level Histograms, IEEE Transactions on Systems, Man & Cybernetics 9 (1), 1979, pp. 62–66.

# Dual Sourcing in Supply Chain Design: A Multi-Dimensional Framework of Risk

**Mario NORBIS**
**Management Department, Quinnipiac University**
**Hamden, CT 06518 USA**

**and**

**Mary J. MEIXELL**
**Management Department, Quinnipiac University**
**Hamden, CT 06518 USA**

## ABSTRACT

Sourcing decisions are a key factor in supply chain security. In this research the authors posit that there are multiple risk types involved in the sourcing decision, and develop a model framework for analyzing these risk components. Specifically, increasing the number of sources minimizes risk related to supply disruptions, but intensifies risk related to security failures. This paper, which is a part of an ongoing research effort that develops a supply chain approach for international logistics, concludes that the best structure for single or dual sourcing from a risk minimization perspective depends on whether the suppliers operate in series or in parallel, and if in parallel, as a backup or as an alternative with a prescribed percentage.

**Keywords:** Supply Chain Security, Risk Analysis, Supplier Selection, Dual Sourcing, Supply Chain Management

## INTRODUCTION

Sourcing decisions have become so important in the life of a corporation to the extent that they can define the corporation its growth and the firm's competitive advantage [1]. The importance of these decisions is underscored in the supply chain. Dependency on suppliers is such that even for service industries half of their purchased services are from other companies [2]. In a supply chain, where firms are linked through bidirectional flows of materials, information, and monetary resources, any change occurring at an upstream stage of the supply chain may affect the performance of firms downstream, through changes in quality, price or availability of material [3].

The work by Anton and Yao [4] concluded that under complete information and with suppliers that have strictly convex costs, a buyer is better off selecting a single source. Since that time, numerous works have analyzed single versus multiple sourcing, including Inderst [5], Zeng [1], Szwejczwski [6], Li [3], Mishra [7] and Jayaraman [8]. Some of these studies include a practitioner's point of view as they focus on the effects that the multiple source has on price along with preventing disruptions in the supply line. Berger [9] analyzes the effect that multiple sourcing may have on risks for the buyer and for the whole supply chain.

In this research the authors start with the premise that there are multiple risk types involved in sourcing. While increasing the number of sources minimizes risks related to supply disruptions, security risks may increase with the inclusion of more members in the supply chain. To explore this premise we build a model that analyzes the effects on security as a buyer moves from single sourcing to dual sourcing.

This introduction is followed by a literature review of published work in the area. The quantification of supplier security, a problem description, and model development are then presented. We introduce a case example and then close the paper with conclusions and ideas for next steps.

## LITERATURE REVIEW

Research that compares single and multiple sourcing procurement includes both theoretical and empirical studies. Some of this research is academically focused while others are more practitioner-oriented.

Our review starts with the contributions by Anton and Yao [10] [4] in which second sourcing and split awards were analyzed in comparison with single sourcing scenarios. Under the conditions analyzed in these cases, single sourcing was preferred. The results from Anton and Yao [4] were extended by Inderst [5], allowing multiple buyers and concluding that the optimality of single sourcing depends on the relative size of the buyer.

Surveys by Szwejczewski [6] and by Larson [11] provide evidence that single-source suppliers deliver higher quality at a lower cost to the buyer. Single-sourcing is also credited with improvement in quality and better pricing by Tullous and Utecht [12] and by Christopher [13].

A descriptive work by Zeng [1] synthesizes available sourcing alternatives into 4 categories: single sourcing, multiple sourcing, single/dual hybrid network, and global sourcing. Single sourcing is understood as using only one source of supply for a particular material or component [14] [15], while multiple sourcing is simply using more than one source. It is argued [16] that risk is increased by single sourcing and that emergency plans in place are needed for emergencies. The existence of little empirical research on the benefits of single sourcing is highlighted by Seitz [17]. The impact of disruptions is graphically analyzed by Li [3] and quantified from the perspective of time and cost.

The single/dual hybrid network involves the use of two or more sources per product type (e.g. an instrument panel for a specific model), with only one source being used for a single component number (e.g. that might specify color). Network sourcing is often the choice of assembly type manufacturing who use large number of parts, as in the auto industry [18]. Mishra [7] shows through an analytic model that order splitting is often the underlying reason for the improvement attributed to dual sourcing. Global sourcing may be defined in this context as the efficient use of worldwide resources [19]. In a highly interconnected world, it also can be argue that all sourcing is global.

Since the events of 9/11, risk has become an important concern for management along the supply chain - in some cases, risk is prioritized over price or quality [20]. For these firms, the "make or buy" decisions previously based on "cost-benefit" analysis are now based on risk-value analysis.

Pope [21] identifies four dimensions of supply chain security: security of the product, security of the information, security of the money and security of the logistics system. The first of these four dimensions pertains exclusively to the supplier. The relevance of this topic is further emphasized by Sheu [22] who concludes that international supply chain security has many issues to resolve to become a fully collaborative system.

Trust is defined by Laeequddin and Sahay [23] as a supply chain partner's threshold level of risk-bearing capacity. Laeequddin *et al.* [24] state that traditionally the intended objective of a supply chain relationship has been to achieve lower cost and to improve product/service delivery. Lately, however, it has been argued that building trust through partnership to better manage risk is now at the heart of the supply chain objectives [25].

Another important topic pertains to security dimensions associated with the supplier selection. Closs and McGarrell [26] propose separate categorization schemes for security themes in suppliers and carriers. This scheme was utilized by Meixell and Norbis [27] when modeling the supplier and carrier choice in the supply chain to improve security as well as by Voss *et al.* [28] when analyzing the security in food supplier selection.

In this context, our research aligns with that of Sheffi [29] in looking at security as a double challenge. We also follow Closs and McGarrell [26] and Voss *et al.* [28] in addressing how to best assign security scores to suppliers. This work also contributes to the literature with a structure for the analysis of risk contributors in the environment of dual sourcing in the supply chain.

## SUPPLIER SECURITY

This research starts with the premise that there is more than one type of risk involved in sourcing decisions and while increasing number of sources minimizes risks related to supply disruptions, security risks may increased with the inclusion of more members in the supply chain. The first type of risk is represented by disruption in the supply chain, motivated by a supplier's inability to provide the right amount of products and services at the required time and with the expected quality. We argue here that this type of risk is a function of the traditional supplier selection criteria: product quality, price and delivery reliability [28]. Here, we use Zeng [1] for the categories in each of these themes as shown in Table 1.

**Table 1. Supplier Security Type 1 Rating Illustration**

| | **Supplier 1 Score** | **Supplier 2 Score** | **Supplier 3 Score** |
|---|---|---|---|
| **Product Quality** | 2 (67%) | 2 (67%) | 0 (0%) |
| Specifications | 3 | 3 | 0 |
| Equipment Capability | 1 | 1 | 1 |
| Quality Assurance Process | 3 | 3 | 2 |
| **Product Price** | 2 (67%) | 3 (100%) | -1 (0%) |
| Cost Structure | 3 | 3 | 0 |
| Financial Capability | 1 | 3 | 1 |
| Supplier Value Analysis Effort | 3 | 3 | 0 |
| **Delivery Reliability** | 2 (100%) | 2 (100%) | -1 (0%) |
| Production Schedule | 3 | 3 | 1 |
| Contract Performance | 3 | 3 | 0 |

The second type of risk is represented by the possibility of a terrorism in any of its forms (i.e. unauthorized people, contraband or weapons of mass

destruction from entering the country), including attack on vessels or on a vessel's contents. Firms can help protect against the risk of terrorist events when selecting suppliers by including criteria for how well suppliers manage supply chain relationships, incorporate security efforts into their existing processes, and develop a process for managing security related incidents. The first theme, relationships, is based on the notion that a supplier that builds and maintains strong relationships with customers, suppliers, government, carriers, and port operators will bring a higher level of security to the venture than one who doesn't. The second theme addresses the degree to which security considerations have been incorporated into existing processes and include, for example, the thoroughness of background checks when screening personnel at the time of hiring and training programs on security practices. These suppliers obviously present a lower security risk than those who do not utilize such practices. The third theme addresses incident security, and the degree to which a supplier has set up management processes for security planning, mitigation, detection, response, and recovery. Planning for response to security incidents puts both the supplier as well as the buyer in better position to react in the event an incident occurs. These themes are further discussed in Closs and McGarrell [26].

The scores illustrated in Table 1 are summed across multiple items within a category, using a framework in which a firm's security practices may be rated as non-existing (0), basic (1), enhanced (2), or advanced (3). When implemented, the values used for any of these indicators could be based on observations made by a buyer within the firm, or perhaps an independent rating agency could support the assessment process by collecting this information for suppliers and making it available to a firms potentially interested in purchasing from the supplier. This framework is further described in Meixell and Norbis [27].

## PROBLEM DESCRIPTION

The problem at hand can be defined as, given more than one supplier with the ability of providing the supply chain each of which can provide the same product or service with different qualification for price, product quality and on time delivery and different security qualifications, what is a suitable methodology to find the best combination to minimize the risk of supply chain disruption as well as threat to the supply chain overall security.

The authors argue that risk of a catastrophic event even when it also represents an interruption of the supply is different in nature from the interruption due to poor quality of the product or delays from the supplier. They rather represent an externally forced interruption by humans that includes damage and may take the form of terrorism, sabotage or other criminal action.

We address this problem by modeling the alternatives and providing case examples of their use.

## MODEL

Let's consider the case of 2 potential suppliers identified as S1 and S2, each able to provide the same product or service with different qualification for price, product quality and on-time delivery. Risk is defined here, consistent with Hutchins [20], as the probability that an event or action adversely affects an organization or its associated supply chain. The probability that each supplier provides the desired product in the amount requested, in time and undamaged is represented by $p_1$ and $p_2$. The risk of service interruption or partial delivery represented by $r_1$ and $r_2$ are calculated as

$$r_1 = 1 - p_1 \quad \text{and} \quad r_2 = 1 - p_2$$

$p_1$ and $p_2$ are evaluated for each supplier based in the characteristics provided by Zeng [1] and represented in Table 1. Also $v_1$ and $v_2$ represent the proportions in which demand is split between the two suppliers in the case parallel splitting demand.

$$v_1 + v_2 = 1$$

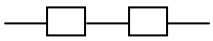
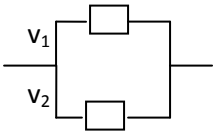
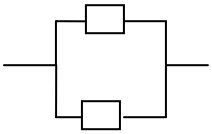
Beyond the ability of the supplier to provide the product or service on time and in the quantity and quality required, the security measures to handle the cargo have become of vital importance [30], [31]. The latter includes security in the facility, government regulation, supplier practices and others as indicated by Closs and McGarrell [26].

The risk or the probability of these events is difficult to evaluate because of lack or irrelevance of historical data [30] and belong more in the area of prediction rather than forecast. This risk is represented as $t_1$ and $t_2$ which are a function on the security scores (SS1 and SS2). It would be expected that $t_1$ and $t_2$ are much smaller than $r_1$ and $r_2$; we can also assume for the sake of discussion and without loss of generality, that

$$t_1 < t_2 \quad \text{and } r_1 < r_2$$

We discuss three different scenarios for the arrangement of suppliers, one in series and two in parallel. For each scenario two alternatives are considered, single supplier or two suppliers [32], [33]. These scenarios and alternatives are summarized in Table 2.

**Table 2. Risk for the 3 Different Supplier Scenarios**

| | Graphical Representation | One Supplier $S_1$ | Two Suppliers $S_1$ and $S_2$ |
|---|---|---|---|
| **Series** |  | | |
| Risk of interruption | | $r_1$ | $r_1 + r_2 - r_1 * r_2$ |
| Risk of catastrophic | | $t_1$ | $t_1 + t_2$ |
| **Parallel Split** |   | | |
| Risk of interruption | | $r_1$ | $v_1 r_1 + v_2 r_2$ <br> $v_1 + v_2 = 1$ |
| Risk of catastrophic | | $t_1$ | $t_1 + t_2$ |
| **Parallel Backup** |  | | |
| Risk of interruption | | $r_1$ | $r_1 * r_2$ |
| Risk of catastrophic | | $t_1$ | Max $\{t_1 , t_2\}$ |

For the series scenario, two independent suppliers in series are continuously and simultaneously active. The risk of undersupply in this case is

$$r_1 + r_2 - r_1 * r_2$$

while the risk of a catastrophic event is

$$t_1 + t_2$$

because any of the two can independently at any time be the origin of a catastrophic threat. In the first parallel scenario, both suppliers are active simultaneously, splitting demand in every cycle. In this parallel-splitting scenario the probability of undersupply becomes

$$v_1 r_1 + v_2 r_2$$

and the probability of fully stopping supply becomes

$$r_1 * r_2$$

Likewise the risk of a catastrophic event in every cycle is

$$t_1 + t_2$$

for the same reasons as in the series scenario.

In the second parallel scenario, suppliers in parallel, one serves as a backup when the first supplier fails to deliver. For this case the risk of undersupply is given by the probability of both suppliers failing

$$r_1 * r_2$$

which is also the risk of total stop of supply. Meanwhile, in any cycle the risk of a catastrophic event is given by the probably of any of the two suppliers independently and not simultaneously providing a threat, which in worst case scenario will be

$$\max (t_1 , t_2)$$

For the three scenarios and the two alternatives, single and dual suppliers, the two types of risks are tabulated in Table 2.

## ANALYSIS OF RESULTS

The results in Table 2 are analyzed for the three scenarios. For the series scenario Table 2 indicates that if a second supplier is added to the more reliable one, the risk of interruption increases as

$$r_1 + r_2 - r_1 * r_2 \geq r_1$$

and the risk of catastrophic event also increases as.

$$t_1 + t_2 \geq t_1$$

If a second source is utilized in parallel and the demand is split in every cycle so that both suppliers are providing part of the demand simultaneously, the risk of under supply increases as

$$v_1 r_1 + v_2 r_2 \geq r_1$$

and the risk of catastrophic event also increases because

$$t_1 + t_2 \geq t_1$$

Finally if a second supplier is added as a backup of the best one in a way that only one is active at a time, the risk of undersupply decreases because

$$r_1 * r_2 \leq r_1$$

while the risk of a catastrophic event increases because

$$\text{Max } \{t_1 , t_2\} \geq t_1$$

generating a tradeoff situation.

## EXAMPLE

A numerical analysis is included in Table 3 below as an illustration for the parallel backup case.

**Table 3. Illustration**

| | Case 1 | Case 2 | Case 3 | Case 4 |
|---|---|---|---|---|
| $r_1$ | .18 | .12 | .09 | .26 |
| $r_2$ | .21 | .18 | .12 | .30 |
| $t_1$ | .0002 | .0001 | .0002 | .0003 |
| $t_2$ | .0003 | .0002 | .0005 | .0006 |
| Risk of under supply | .038 | 0.0216 | .011 | 0.078 |
| Catastrophic risk | .0003 | .0002 | .0005 | .0006 |

## CONCLUSIONS AND NEXT STEPS

In this research we propose that there are multiple risk types involved in a supplier sourcing decision. Specifically, increasing the number of sources minimizes risk related to supply disruptions, but intensifies risk related to security failures. We develop a model framework for analyzing these risk components and show that the best structure for single or dual sourcing from a risk minimization perspective depends on whether the suppliers operate in series or in parallel, and if in parallel, as a backup or as an alternative with a prescribed percentage.

The next steps in this research are to develop an approach for estimating $t_1$ and $t_2$, given supplier security scores from the risk assessment matrix along with other related information in the supply chain. Also, the expected loss may be factored into the analysis, so that supplier selection reflects the overall impact along with the probability of a negative event occurring.

## REFERENCES


[1] A. Zeng, "A Synthetic Study of Sourcing Strategies", **Industrial Management + Data Systems ,** 2000, Vol. 100, 5, pp. 219-229.

[2] M. Mihaly, "Strategic Sourcing", **Industry Week**, 1999, Vol. 248, No. 4, pp. 3-5.

[3] G. Li, Y. Lin, S. Wang, H. Yan, "Enhancing Agility by Timely Sharing of Supply Information," **Supply Chain Management : An International Journal**, 2006, Vol. 11, No. 5, 425-435.

[4] J. Anton, D. Yao, "Split Awards, Procurement and Innovation", **RAND Journal of Economics**, 1989, Vol. 20, pp. 538-552.

[5] R. Inderst, "Single Sourcing versus Multiple Sourcing", **The RAND Journal of Economics, 2008,** Vol. 39, No. 1, Spring 2008, pp. 199-213.

[6] M. Szwejczwski, K. Goffin, F. Lemke, R. Pfeiffer, B. Lohmuller, "Supplier Management in German Manufacturing Companies, An Empirical Investigation", **International Journal of Physical Distribution & Logistic Management,** 2001, Vol. 31, 5, pp. 354-373.

[7] A. Mishra, P. Tadikamalla, "Order Splitting in Single Sourcing with Schedule-Release Orders", **Journal of the Operational Research Society,** 2006, Vol. 57, pp.177-189.

[8] V. Jayaraman, R. Srivastava, W. Benton, "Supplier Selection and Order Quantity Allocation: A Comprehensive Model", **Journal of Supply Chain Management**, 1999, Spring 1999, Vol. 35, No. 2, pp. 50-58.

[9] P. Berger, A. Zeng, "Single versus Multiple Sourcing in the Presence of Risk, **Journal of the Operational Research Society**, 2006, Vol. 57, 250-261.

[10] J. Anton, D. Yao, "Second Sourcing and the Experience Curve: Price Competition in Defense Procurement", **RAND Journal of Economics**, 1987, Vol. 18, pp. 57-76.

[11] P. Larson, J. Kulchisky, "Single Sourcing and Supplier Certification: Performance and Relationship Implications", **Industrial Marketing Management** 1998, Vol. 27, pp. 73-81.

[12] R. Tullous, R. Utecht, R., "Multiple or Single Sourcing?", **The Journal of Business & Industrial Marketing,** 1992, Vol. 7, No. 3, pp. 5-17.

[13] M. Christopher, "The Agile Supply Chain: Competing in Volatile markets", **Industrial**

**Marketing Management** 1998, Vol. 29, pp. 37-44.

[14] R. Newman, "Single Source Qualifications", **International Journal of Purchasing and Material Management**, 1988; Vol. , pp. 9-19.

[15] J. Ramsay, I. Wilson, "Sourcing / Contracting Strategy Selection", **International Journal of Operation & Production Management,** 1990, Vol. 10, No. 8, pp. 19-20.

[16] K. Natchweh, "Notfallmanagement als Teil des Lieferantenmanagements", **Beschaffung Aktuell**. 1999, Vol. , No. 8, pp. 40-43.

[17] H. Seitz, "Neue Wege in der Wareneingangsprüfung**", QZ Zeitschrift Für industrielle Qualitätssicherung**, 1991, Vol. 36, No. 6, pp. 330-3

[18] P. Hines, "Network Sourcing: a Hybrid Approach", **International Journal of Purchasing and Material Management,** 1995; pp. 18-24.

[19] R. Spekman, "US Buyers' Relationship with Pacific Rim Sellers", **International Journal of Purchasing and Material Management**, 1991, Vol. , No. , pp. 2-10.

[20] G. Hutchins, "Risk Management in the Supply Chain, Proceedings", **Annual Quality Congress, ASQ**, 2003, pp. 49-58.

[21] J. Pope, "Dimensions of Supply Chain Security", **Southern Business Review**, Summer 2008.

[22] C. Sheu, L. Lee, B. Niehoff, "A Voluntary Logistics Security Program and International Supply Chain Partnership", **Supply Chain Management : An International Journal**, 2006, Vol. 11, No. 4, pp. 363-374.

[23] M. Laeequddin, V. Sahay, "Risk the Limiting Factor of Supply Chain Partner Relationship Trust: A Conceptual Frame Work", **Decision Sciences and Technology for Globalization,** Allied Publishers, New Delhi, pp. 3-17. 2008.

[24] M. Laeequddin, G. Sardana, B. Sahay, A. Waheed, V. Sahay, "Supply Chain Partners' Trust Building Process Through Risk Evaluation: the Perspectives of UAE Packaged Food Industry", **Supply Chain Management : An International Journal,** 2009, Vol. 14, No. 4, 425-435.

[25] R. Spekman, E. Davis, "Risky Business: Expanding the Discussions on Risk and the Extended Enterprise", **International Journal of Physical Distribution & Logistics Management,** 2004, Vol. 34, No. 5, pp. 414-33.

[26] D. Closs, E. McGarrell, "Enhancing Security Throughout the Supply Chain", **IBM Center for the Business of Government**. 2004.

[27] M. Meixell, M. Norbis, "Integrating Carrier Selection with Supplier Selection Decisions to Improve Supply Chain Security", Working paper. 2010

[28] M. Voss, D. Closs, R. Calantore, O. Helferich, C. Speier, "The Role of Security in the Food Supplier Selection Decision", **Journal of Business Logistics**, 2009, Vol. 30, No. 1

[29] Y. Sheffi, "Supply Chain Management under the Threat of International Terrorism", **International Journal of Logistics Management**, 2001, Vol. 12, No. 2, pp. 1-11

[30] K. Bichou, A. Evans, "Maritime Security and Regulatory Risk-Based Models: Review and Critical Analysis" Chapter 16 in "Risk Management in Prot Operations, Logistics and Supply-Chain Security" Informa Ed. London 2007, pp. 265-279.

[31] M. Meixell, M. Norbis, "A Review of the Transportation Mode Choice and Carrier Selection Literature", **International Journal of Logistics Management,** 2008, vol 19, no. 2, pp. 183-211.

[32] L. Leemis, "Reliability: Probabilistic Models and Statistical Methods". **Englewood Cliffs**, New Jersey, Prentice-Hall. (1995).

[33] M. Meixell, "Quantifying the Value of Web-Services in Supplier Networks**." Industrial Management & Data Systems,** 2006, Vol. 106, No. 3, pp. 407-422.

# A Key Management Solution for Secure Routing in Mobile Ad Hoc Networks

**Sulaiman ASHRAPH**
asulaiman@nur.ac.rw
**National University of Rwanda**
**Butare, Rwanda**

**and**

**Dawoud S DAWOUD**
dshenouda@nur.ac.rw
**National University of Rwanda**
**Butare, Rwanda**

**and**

**Adronis NIYONKURU**
aniyonkuru@nur.ac.rw
**National University of Rwanda**
**Butare, Rwanda**

## ABSTRACT

This paper investigates the need for security and more specifically key management for secure routing in mobile ad hoc networks. A key management protocol is proposed for on-demand ad hoc routing protocols. The proposal provides key distribution and security evaluation to allow the most secure path to be selected during route discovery. Simulations modeled in ns-2 investigate the success of the key distribution mechanism and security scheme.

**Keywords**: Ad hoc networks, Trust chain, On-demand routing protocol, Secure Routing.

## 1. INTRODUCTION

With recent wireless technology advances mobile ad hoc networks have found increasing application in the military and commercial domain. However the unique characteristics of mobile ad hoc network make them difficult to secure. Such characteristics include the lack of network infrastructure, no prior relationships, unreliable multi-hop communication channels, resource limitation and node mobility. Before these dynamic networks can be deployed into the commercial and military arena they must be secured against malicious attackers.

We identify that most of the security attacks target the network layer and more specifically the routing protocol. Such attacks include blackhole attacks, wormhole attacks, eavesdropping attacks, byzantine attacks, resource consumption attacks and routing table poisoning. The network layer provides a critical service to the mobile ad hoc network, called the routing protocol. In the context of trust and security the provision of secure routes is one of the most vital elements for trust establishment. Adversaries target the routing protocol specifically and a secure routing solution is needed for a secure implementation and deployment of ad hoc networks.

Secure ad hoc routing protocols exist but their operational assumptions typically include an existing key management system. The following section investigates the operational requirements for existing secure ad hoc routing protocols. Secure ad hoc routing protocols are divided into three categories: symmetric cryptography, asymmetric cryptography, and reputation based protocols.

Although symmetric cryptographic approaches do not rely on a public key infrastructure they still require some kind of key management in an ad hoc network. The SEAD protocol [1] designed for the table-driven DSDV routing protocol requires a key management mechanism to distribute an authenticated initial hash element. SEAD uses a one-way hash function to provide hop-by-hop authentication for routing packets hop count and sequence numbers. Ariadne [2] proposed a DSR based on-demand protocol uses TESLA authentication [3] to provide end-to-end authentication. A key management system is assumed present to distribute the TESLA keys. TESLA authentication also requires clock synchronization between participating nodes which is difficult without the presence of an online TTP.

ARAN [4], SAODV [5] and SLSP [6] use asymmetric cryptography and key management is simply assumed for each of these protocols. ARAN assumes that an online TTP is present that acts as a certificate authority (CA) to provide end-to-end authentication in an on-demand environment. Prior shared secrets are assumed between all participating nodes and the CA. The SAODV protocol is based on the AODV on-demand routing protocol and provides end-to-end authentication by authenticating the routing packets' mutable fields with a hash chain and immutable fields with a digital signature [5]. SAODV assumes the presence of a key management system to distribute keys and initial hash elements. SLSP secures a table driven link state routing protocol (OLSR) by providing secure neighbor discovery and authenticated link state updates [6]. SLSP uses digital signatures and assumes that nodes enter the network with asymmetric key pairs and a key management scheme is present to certify the keys in the network.

ODSBR [7] is based on the on-demand DSR routing protocol. ODSBR authenticates its routing packets with digital signatures

and a public key infrastructure is assumed to manage the keys. ODSBR employs a reputation based mechanism to monitor data packets and maintain a path specific rating list. Shared keys are assumed to allow for authenticated acknowledgement messages to be transmitted between a source and probe nodes which provide the evidence for path rating. CONFIDANT [8] is another reputation based solution which does not use any cryptographic techniques and therefore does not require a key management system. CONFIDANT does assume pre-exiting relationships between a small number of nodes called friends. The CONFIDANT is an on-demand solution which provides a node specific reputation list to help selecting secure paths during route discovery.

The observation is made that most secure ad hoc routing protocols assume the existence of a key management system to certify, authenticate, and distribute keying information. Pure mobile ad hoc networks cannot assume the existence of a trusted third party (TTP) and must address the problem of key management.

## 2. OUR APPROACH

**a. System Model**

We view the trust problem in ad hoc networks as a weighted trust graph *G(V,E)* where the vertex *V* represent nodes and the edges *E* represent paths. There is no TTP present during communication. The routing environment is on-demand. Each node is assumed to have an asymmetric key pair before entering the network.

We design a key management system to provide key management for secure on-demand routing protocols. The system operation is as follows. A key distribution scheme is proposed which is divided into two components: localized key exchanges and remote end-to-end key exchanges. A local authentication is performed by exchanging keys between close proximity neighbours over a location-limited channel. Remote end-to-end key exchange uses the established localized relationships to share keys remotely across multiple hops. The security evidence provided by certificates is allowed to influence the selection of routes during the route discovery phase. We propose a security evaluation metric which aggregates trust along a path based on a security metric and path distance.

Similar to [14], we consider a fully distributed network of wireless nodes with generic medium access control (such as IEEE 802.11) and secure on- demand routing mechanisms. Nodes can be stationary or move with low to high mobility speeds (0*m/s* -20*m/s*). We assume that there are no pre-existing infrastructure and no form of *online* trusted authority to assist the key distribution mechanism. Since we are considering *authority-based* MANETs as defined in [14], there exists an *offline* authority to bootstrap the system; before users join the network they have to acquire a certificate from the *offline* trusted authority. The trusted authority thus only provides each node with its own certificate and not with the certificates of any other nodes. This requirement is fundamental to ensuring scalability and on-demand network formation. Each node is also issued with the authentic public key of the trusted authority and a universal set of system parameters. The certificate must contain the offline authority's identity, the node's public key and identity/network address, a unique sequence number, certificate generation date and expiry date

**b. Key distribution scheme**

**1) Localized key exchange:** Neighbouring nodes can securely exchange keying material without the presence of a TTP. Nodes in close proximity first exchange pre-authentication information across a location-limited channel. Location-limited channels include infrared, physical contact, audio, etc. Because of the nature and characteristics of a location-limited channel like infrared the neighbors can be assured that the pre- authentication is legitimately from the sender. After users exchange pre-authentication information a group certificate exchange scheme is implemented over the main wireless channel. Certificates binding a node's unique ID and public key are exchanged and authenticated using the previously acquired pre- authenticated information.

Following the routing protocol the localized key exchange has two approaches. Firstly following a route request immediate nodes perform a localized key exchange. Alternatively localized key exchange would follow a hello packet broadcast of an on-demand routing protocol like AODV.

**2) Remote end-to-end key exchange:** The certificate of the source node *A* is forwarded along the request route packet *RREQ* independent of the routing packets. At each hop the intermediate node is inquired if it possesses the source's certificate. If not a separate unicast message is sent back to the previous hop requesting the sources certificate. This procedure follows the *RREQ* until it reaches its target. The destination node *B* caches the sources certificate and replies with a route reply message *RREP*. The *RREP* is forwarded along the reverse path and at each intermediate node the local certificate repository is checked if it contains the destination's certificate. If the certificate is not found a unicast certificate exchange similar to the forward path one is performed. All the certificates are verified on the *REPP* route. This approach minimizes unnecessary certificate verifications on paths which will not be selected.

When a node (RN) receives a routing control packet it checks in its certificate database if it has the certificates of the packet originator (ON) and the previous hop node (PN) on the forward route. If RN has both the certificates of ON and PN (*CertON* and *CertPN*), it can process the control packet as normal. If not, it requests both the certificates from PN. If RN does not have the certificate of PN, then it also sends its own certificate with the request to the previous- hop.

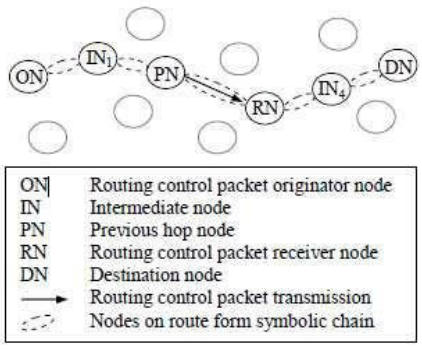



**Figure 1:** Certificate distribution main procedure

Note that if RN is the first-hop on the route, then the previous-hop node and the control packet originator node will be the same entity. The routing messages thus effectively chain nodes together and allow them to relay all keying material, as required, along the virtual chains.

The proposed key distribution scheme, works as follows:

– When any node in the network receives a RCP it first determines if the originator of the message (ON) has the same network address as the previous hop node (PN) on the forward route, that is, RN has to determine if ON is the first-hop.

In case when the addresses of ON and PN are the same, the RN runs:

**Protocol 1:**
RN searches its certificate repository for ON certificate

If found,
Process the routing message RCP
else
ON and RN Exchange Certificate (Peer-to- Peer)
$CertRN \rightarrow ON,\ ON \rightarrow CertON$
Process the routing message RCP

– If the ON address and the previous-hop node (PN) address are not equal, RN runes protocol 2. The RN will search its certificate repository for *CertON and CertPN*.

**Protocol 2:**
RN searches its repository for CertPN
If Found,
Search for CertON
If Found
No action, process routing packet as normal
else
Peer-to-Peer certificate exchange
$CertRN \rightarrow PN,\ CertPN \rightarrow RN$
else
Search for $Cert_{ON}$
If Found
$Cert \rightarrow PN, Cert_{PN} \leftrightarrow RN$
else
Peer-to-Peer certificate exchange
$CertRN \ // \ Cert \rightarrow PN$
$CertPN \ // \ CertON \rightarrow RN$
Process routing packet as normal

**c. Security Evaluation Scheme**

The key distribution scheme follows the route discovery phase exchanging keys locally and remotely. The security evaluation scheme enhanced the route discovery phase. This component's core approach is to provide trust path selection based on both distance and a security metric. The ad hoc network is modeled as a weighted trust graph *G(V,E)* where the edges *E* are nodes with a weighted trust metric *t*. This trust metric is assumed to be available based on either public key certificates or a reputation-based monitoring system. The vertexes *V* represent the weighted paths or routes in the network. We propose an evaluation metric for these weighted paths. We propose the use of the semiring mathematical operators $\oplus$ and $\otimes$ to aggregation trust along a multi-hop path of the ad hoc on-demand protocol. The distance semiring approach uses the operator to aggregate metrics along a path like parallel resistors would be summed ie. $\frac{1}{R_T}=\frac{1}{R_1}+\frac{1}{R_2}+...+\frac{1}{R_n}$. The aggregated value will decrease along a path and the total is always less than the summations smallest component. This description aligns with our design specifications. The aggregated value will be low for routes with high hop counts. Secondly it will reflect the trust characteristics of a trust chain which states that a chain is only as strong as its weakest link.

Following the route discovery phase, at each hop of a *RREQ* the current path weight is calculated and compared to an implicit path revocation threshold. If the aggregated trust is more than the threshold value the path and associated weight are stored in the routing table corresponding to the reverse route and the *RREQ* is forwarded. Similarly at each hop of a *RREP* the current path weight is calculated and stored in the routing table matching the forward route. Multiple replies to a single *RREQ* are filtered by the hop count metric in standard AODV and DSR on- demand routing protocols. We propose that routes are filtered based on their sequence number and their semiring based weighted path metric. This filters the most recent routes and based on the nature of the weighted path metric, hop count and trust are both accounted for.

## 3. RELATED WORK

Key management solutions are presented in [9-11]. The self issuing certificate chain approach in [12] proposes a self organized key management solution that treats the network as a weighted direct graph. This proposal is a proactive methods designed for application layer security.

Proximity based key agreement is investigated in [13] and again by Capkun in [14] where security relies upon nodes' mobility to provide multiple close proximity key exchanges. This proposal is limited by its reliance on mobility.

A proactive reputation-based solution is proposed in [15] which operates on the application layer. Semi- ring mathematics [16] is used to realize the proactive generic-single-source-shortest-distance algorithm. The majority of literature mentioned functions in a proactive manner for application layer solutions. We propose a reactive approach for the network layer

## 4. EVALUATION

Our approach is simulated as a C++ network layer application in ns-2. It is planned for ad hoc on- demand routing protocols and designed on the AODV routing protocol but our solution is not protocol specific. The key distribution scheme is simulated and test in large node environment with varying mobility and load. The packet delivery ratio is examined to investigate how the security mechanism affects the success of the routing protocol. The normalised routing load is investigated to show the overhead of the key exchange components. The security evaluation scheme is difficult to evaluate using simulations. The scenarios is established where the assumed weighted node edges carry a security metric

which identifies fault detection or data transmission errors carried out by nodes for example in [7, 8]. This allows the proposal to protect against black hole attacks. Simulations are set up which investigate the packet delivery ratio with a changing number of adversary nodes.

**i) Performance Metric**

The following quantitative metrics are used to analyze the performance of the routing protocols in mobile ad hoc networks.

**(a) Packet Delivery Ratio:** The packet delivery ratio (PDR) represents the percentage of data packets that are successfully received by their intended destination. This metric is also known as throughput and is considered a measurement of the effectiveness of a routing protocol. The equation for PDR is:

$$PDR\% = \frac{\sum_1^n CBRrec}{\sum_1^n CBRsent} \times 100$$

where $\sum_1^n CBRrec$ and $\sum_1^n CBRrec$ are the number of CBR data packets received and sent respectively.

**(b) Routing Overhead:** A routing protocol uses control packets to establish routes on which data packets are transmitted. Control packets are separate from data packets but share the same communication channel. Due to the lack of channel capacity in mobile ad hoc networks a large number of control packets can result in poor network performance. Key management would require additional control packets to achieve key management functionality this will be reflected in the simulations. The routing overhead is also known as a routing protocol's internal efficiency and will represent the number of control packets used for a given protocol.

**(c) Average End-to-End Delay:** This is a qualitative measurement of the delay of data packets. The average end-to-end delay of a data packet is the time from which it is created at the source and when it arrives at the intended destination. The delay includes propagation and queuing delay. Delay can be caused by a high number of control packets propagating in the network or a high computational overhead for the given protocol. The average end-to-end delay is calculated as follows,

$$End\ to\ End\ Delay = \frac{\sum_1^n (CBRsendtime - CBRrecvtime)}{\sum_1^n CBRrec}$$

where *CBRsendtime* and *CBRrecvtime* represent the record times that a CBR data packet was sent and received

**ii) Simulation Model**

A routing protocol was designed in C++ based on the AODV routing protocol available in the ns-2.31 package. The routing protocol is programmed as a routing agent class. The routing agent handles the establishment of routes and certificate distribution. Modifications are made to the AODV routing agent at the *RecvRequest, SendRequest, RecvReply*, and *SendReply* functions. These modifications allow for the distribution of separate certificate packets, triggered by the routing packets. The routing agent's packet header was modified to include a certificate control packet *CertS*. The size of the certificate included is 450 bytes. The size of the certificate control packets is increased resulting in an effective delay in communication simulating the transfer of actual certificates. A certificate table is included at each node *CertTable* which is updated by certificate control packets. The certificate table is linked to the routing table and each node is responsible for managing its own certificate table.

A simulation *tcl* file is written to setup the mobile ad hoc network's desired simulation scenario, traffic and mobility model. The trace support files in ns-2.31 were modified to support the routing agent allowing the inclusion of certificate control packets and trust information. As a result the output trace and *nam* files reflect the operation of the routing agent. Figure 2 shows a sample output of the *nam* simulation file.

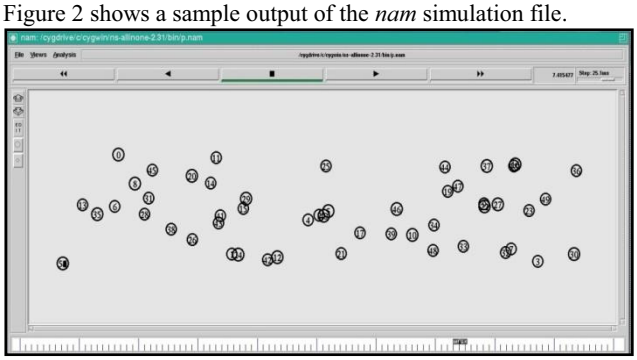


**Figure 2:** Sample *nam* simulation file illustrating typical network topology

**iii) Packet Delivery**

The packet delivery results for the AODV routing protocol is presented in Figure 3 & 4. Figure 3 represents a simulation environment with a pause time of 0 seconds. This represents a network of nodes that are continually moving, while Figure 4 represents a partially stable network. The observation is made that as the speed increases both protocols throughput decreases. At high speeds the network topology changes rapidly causing breakages in routing links. The reduction in packet delivery at high speeds is because both protocols will drop data packets as a result of increased routing breakages. The stable network in Figure 4 shows better performance at higher speeds because the number of route link breakages is reduced as a result of a larger pause time. A large pause time represents a network that will move at a given speed then pause in a fixed location for a set amount of time. During this time routing link breakages are not expected until movement commences again.

A certificate distribution scheme would expect a severe reduction in performance due to an excessive number of packets being transmitted in the network or the additional size of the control packet. A conventional certificate distribution scheme, suggested as a possible solution in [8], simply includes the source certificate in the request packets *RREQ* and the destinations certificate in the reply packets *RREP*. This method was implemented as a separate routing agent *AODVcert* in ns2. A similar method is suggested in [6]. Implementation includes increasing the packet size of the routing control packets to include a 450 byte certificate. This effectively increased the regular 56 byte AODV route control packets to 506 bytes. Such an approach would result in the simplest method of certificate distribution but transmitting 450 bytes more data per control packet would severely reduce the network performance.

## 5. CONCLUSION

It is concluded from investigation of existing secure ad hoc routing protocols that the majority assume the existence of a key

management system to distribute and manage the asymmetric or symmetric keys. We propose a key management solution that provides a key exchange scheme and a security evaluation scheme. The proposal is design for an on-demand ad hoc routing protocol.

Simulations in ns-2 investigate the packet delivery ratio of the solution's key exchange scheme compared at varying load and mobility. The security evaluation scheme allows for the most trusted route to be selected during the route discovery phase. An available security metric is assumed and simulations investigate the success of preventing black hole attacks.

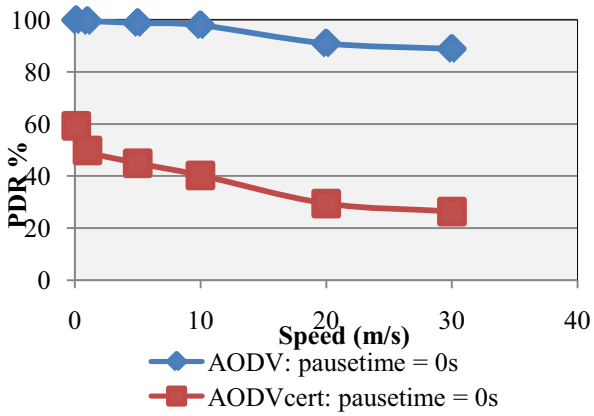



**Figure 3:** Packet Delivery Ratio for highly mobile network (0 second pause time)

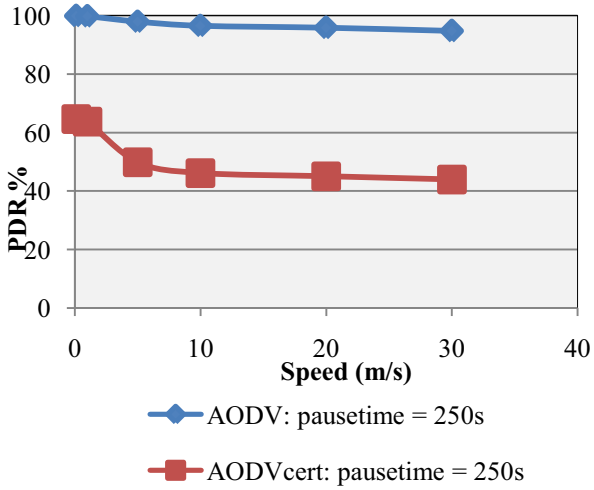



**Figure 4:** Packet Delivery Ratio for partially stable network (250 second pause time)

## 5. REFERENCES


[1] Y.-C. Hu, D. B. Johnson, and A. Perrig, "SEAD: Secure Efficient Distance Vector Routing for Mobile Wireless Ad Hoc Networks", in ***Proceedings of the Fourth IEEE Workshop on Mobile Computing Systems and Applications*: IEEE Computer Society**, 2002.

[2] Y.-C. Hu, A. Perrig, and D. B. Johnson, "Ariadne: a secure on-demand routing protocol for ad hoc networks", ***Wireless Networks***, vol. 11, pp. 21-38, 2005.

[3] A. Perrig, R. Canetti, D. Song, and D. Tygar, "Efficient and Secure Source Authentication for Multicast", **Network and Distributed System Security Symposium (NDSS'01)**, 2001.

[4] K. Sanzgiri, B. Dahill, B. N. Levine, C. Shields, and E. M. Belding-Royer, "A Secure Routing Protocol for Ad Hoc Networks" in ***Proceedings of the 10th IEEE International Conference on Network Protocols*: IEEE Computer Society**, 2002.

[5] M. G. Zapata, "Secure ad hoc on-demand distance vector routing," ***SIGMOBILE Mob. Comput. Commun. Rev.***, vol. 6, pp. 106-107, 2002.

[6] P. Papadimitratos and Z. J. Haas, "Secure Link State Routing for Mobile Ad Hoc Networks," in ***Proceedings of the 2003 Symposium on Applications and the Internet Workshops (SAINT'03 Workshops)*: IEEE Computer Society**, 2003.

[7] B. Awerbuch, R. Curtmola, D. Holmer, C.Nita-Rotaru, and H. Rubens, "ODSBR: An on-demand secure Byzantine resilient routing protocol for wireless ad hoc networks," ***ACM Trans. Inf. Syst. Secur.***, vol. 10, pp. 1-35, 2008.

[8] S. Buchegger and J.-Y. L. Boudec, "Performance analysis of the CONFIDANT protocol," in ***Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing,*** Lausanne, Switzerland: ACM, 2002.

[9] S. Basagni, K. Herrin, D. Bruschi, and E. Rosti, "Secure pebblenets," in ***Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing,*** Long Beach, CA, USA: ACM,2001.

[10] L. Zhou and Z. J. Haas, "Securing Ad Hoc Networks," ***IEEE Network: special issue on network security***, vol. 13, pp. 24-30, 1999.

[11] D. B. Smetters, D. Balfanz, D. K. Smetters, P. Stewart, and H. C. Wong, "**Talking To Strangers: Authentication in Ad-Hoc Wireless Networks**," 2002.

[12] S. Capkun, L. Butty, and J.-P. Hubaux, "Self-Organized Public-Key Management for Mobile Ad Hoc Networks," ***IEEE Transactions on Mobile Computing***, vol. 2, pp. 52-64, 2003.

[13] A. Scannell, A. Varshavsky, A. LaMarca, and E. D. Lara, "Proximity-based authentication of mobile devices," ***International Journal Secure Networks***, vol. 4, pp. 4-16, 2009.

[14] S. Capkun, L. Buttyan, and J.-P. Hubaux, "Mobility Helps Peer-to-Peer Security," ***IEEE Transactions on Mobile Computing***, vol. 5, pp. 43-51, 2006.

[15] G. Theodorakopoulos and J. S. Baras, "On Trust Models and Trust Evaluation Metrics for Ad-Hoc Networks," ***IEEE Journal on Selected Areas in Communications***, vol. 24, pp. 318-328, 2006 2006.

[16] M. Mohri, "Semiring frameworks and algorithms for shortest-distance problems," ***Journal on Autom. Lang. Comb.***, vol. 7, pp. 321-350,2002.

# Use of Anonymous Verifiable Public Key Certificates for Vehicular Communication

**Dawoud S DAWOUD**
dshenouda@nur.ac.rw
**National University of Rwanda**
**Butare, Rwanda**

**and**

**Peter D. Dawoud**
pdawoud@gmail.com
**University of Waterloo**
**Canada**

## ABSTRACT

The paper describes a practical architecture and protocols for offering privacy for a user who uses certificates containing pseudonym, while retaining the ability to map such a certificate to the real ID of the user requested it.

The technique used avoids some of the shortages in the current systems that use pseudonyms. A split Trusted Authority (TA) model is used. The system allows issuing the same key for more than one user that matter that strengthens the anonymity.

**Keywords**: Vehicular Communication, Blind Signature, Public Key Infrastructure, Pseudonymous, Anonymous Certificates.

## 1. INTRODUCTION

With the wide spread of on-line services, it is required from the users to prove their identities to the service provider before gaining an access to the services. This takes place through process called "authentication", which is commonly performed through the use of logon credentials such as a user password and ID. The user is considered authenticated if he has knowledge of the credentials. One weakness of using only logon credentials is that they are susceptible to being accidentally revealed, lost, or stolen. As a result, some services may require a more in-depth authentication process that relies upon a digital certificate that is issued and verified by a certificate authority (CA) using what is known as public key infrastructure ("PKI") security model which relies upon asymmetric (i.e., public key) cryptography. Under PKI, a private key is used for digital signing and authentication, and the certificates are used to distribute a public key portion of a PKI key pair.

But while a unique certificate can provide evidence that a user has the right to access an on-line service, under current PKI methodologies, the certificate is vulnerable to being used to improperly identify and/or track the user. In addition, third parties may attempt to gain access to a private key and intercept information intended for a user to thereby compromise the integrity of the PKI security model.

In vehicular communication (VC), authentication is needed when the car is asking for an on-line service and also when the vehicle is trying to offer a service; sending safety messages to other vehicles on the road. In the last case, any vehicle before accepting the safety message must be sure from the key given to decrypt the message and more important that it is possible to bind this message to the sending vehicle in case if the message contains faulty information, i.e. the sending vehicle must authenticate itself.

In VC as well as many of the on-line services, when users are using certificates to access resources, there are legitimate concerns about personal privacy, and thus there are increasing demands for privacy enhancing techniques on the public Internet as well as in VC. In other words, the anonymity of a user, whither a vehicle driver or a user asking for service from a provider, is required to be preserved when authenticated (note: from now on we are using the words user, driver and vehicle alternatively). Anonymous certificate is one that its subject field does not indicate the holder's true identity. Such a certificate is called "pseudonymous" since an X.509 certificate must contain an ID.
For accountability, the method used to achieve anonymity must guarantee traceability; binding the holder of the anonymous authenticating certificate to his real ID. In other words, it is possible to map an anonymous certificate to the one to whom it is issued. Without such mapping, the user might abuse the anonymity offered to him by the certificate.

This can be achieved through the use of an anonymous and verifiable (i.e., "blind") certificates. In vehicular communication (VC) as well as in many other applications, they use pseudonyms [1-2] as the "blind" signature that binds some cryptographic material (normally the public key of the user/vehicle) to the real ID of the user.

The pseudonyms are normally issued during registration, case of VC, or when a user applies for subscription for services. This can take place on-line. Normally, the AC issues a set of pseudonyms for the user and when the user exhausts this set, he can apply on-line to issue another set.

Each pseudonym is binding the real ID of the user to one of a set of public keys. This set of public keys may be proposed by the user or can be allocated to him by CA. When a vehicle sends a safety message to other vehicles on the road, the sending vehicle selects one pair of the certified keys, encrypts the message with the private key and sends the corresponding public key and pseudonym with the message. The vehicles are using the public key to decrypt the received encrypted message. In case of requesting a service from a provider, the user selects one pair of the certified keys, sends the message to the provider who will

verify the certificate, and uses the key to encrypt the material that the user requested.

If the user uses the pseudonym more than one time, the public key can be used by a third party to track the user (vehicle), a fact that contradicts the anonymity [4]. To avoid that, in such a system, the public key must be used once and the drivers must switch from one pseudonym to another at certain zones [3]. The use of the key once to avoid tracking the car, results in the possibility of exhausting the initially issued certificates during registration or subscription, and, accordingly, the user has to contact the CA to issue more pseudonyms.

In this paper we are using a technique that allows more than one user to use the same public key. This is another difference with X.509 certificate [5]; X.905 certificate must contain the ID of the holder and CA must not issue multiple certificates with the same subject name (the key in our case) to different entities (vehicles or generally users). In our case, the certificate does not contain any information that directly refer to the real ID of the user, and it is allowed to have more than one certificate for the same subject; the public key. The certificate in this case must be distinguished from conventional pseudonyms; it is better to call it anonymous certificate. The use of this technique, as we will see, allows high anonymity and at the same time cancels the need for sending to CA to issue more pseudonyms.

At first the paper will describe a separation-of-authority model that will be used to issue traceable anonymous certificates. The means to counter a claim by malicious (dishonest) user who later denies sending a message will be presented; upon the agreement between the authorities who issued the anonymous certificate, it is possible to map the anonymous certificate to the real identity of the user. The means to stop the dishonest driver or the user, who, for example, did not fulfil his agreement with the service provider, are also discussed. The idea of using the same key pair by more than one user will be discussed. How this technique supports the functions of traceability and revocation will also be discussed.

## 2. PKI MODEL USED

To avoid the possibility of "big brother" attack, in our paper we are using the traditional PKI model that separates the functions between Registration Authority (RA) and Certificate Authority (CA). This, as a matter of fact, goes with the real environment of VC where the vehicles have to go to RA for registration or renewing the registration and the CA is the one responsible for issuing the pseudonyms (anonymous certificates). The model is also common in e-voting. RA and CA form a Trusted Authority (TA). The RA in this model acts as a Blind Issuer (BI) and the CA acts as an Anonymity Issuer (AI). The AI has knowledge on the issued certificates but has no knowledge of the user real identity. The BI knows the user's real identity but has no knowledge of the issued certificates. The conditional traceability is possible only with the cooperation of AI and BI. Normally, RA (BI) and CA (AI) are running by different organizations and managed by different staff. This gives higher confidence in the anonymity of the issued certificates.

In the split-authority model, BI and AI collaborate to generate a public key pair ($SK_{CA1}$, $PK_{CA1}$) using threshold signature scheme, e.g. DSA-based threshold signature scheme [6][7]. The private key is divided into exactly two equal shares, one held by BI and the other half by AI. AI and BI use their parts to create signatures and the public key will be published as the public key of Trusted Authority (TA). AI in this model has another pair of keys ($SK_{CA2}$, $PK_{CA2}$) known only for him.

The PKI environment in the model consists of: A Trusted Authority (TA) consists of the Registration Authority RA that acts as BI and CA that acts as AI; Roadside Units (RSU), which are fixed unites distributed all over the network; Services Providers (SP) provide on-line services; On-Board Units (OBUs) (Sometimes called Hardware Secure Devices (HSDs)). The OBUs are embedded in the vehicles. OBUs can communicate with other OBUs through V2V communication, with RSUs through V2I, or with the providers through V2P communications; and a fixed size set of PKI key pairs. OBUs are normally reconfigurable such that they can be reconfigured by downloading data from the CA.

The two entities forming the TA, BI and AI, are configured as part of PKI infrastructure implementation in the on-line environment that uses a system of digital certificates to verify and authenticate the validity of a user that is involved in a transaction through the use of public key cryptography. The transaction can be asking for a service from a provider or offering a service (e.g. safety message) for other users (vehicles).

In this model, the BI (RA), as it is normal, interacts with the user (e.g., vehicle's owner) to verify his real identity $ID_v$. BI maintains a database that can be used to map an anonymous certificate to the user to whom it was issued, but only with the cooperation of the AI. This mapping will be initiated normally upon request of a legal entity if there is evidence that the user to whom the certificate was issued abused the anonymity provided by the blind certificate. This can be initiated by a service provider or, in case of vehicular communication, if a number of vehicles agreed that a driver of a car is not honest and he is issuing dishonest safety messages.

BI after authenticating the user supplies him with some cryptographic material to let the user be able to send a request to AI to issue the anonymous certificates. In case of vehicular communication, it is also possible that BI, on behalf of the vehicle, may send the request message to AI. BI attach with the request some cryptographic material and information. The cryptographic materials, in the two cases, are needed to prove for AI that the user is ready authenticated itself to BI. This is to guarantee that it is possible to map the anonymous certificate to the real ID of the user to whom it is issued. The cryptographic material supplied by BI will also be used by AI as a database entry key for the set of certificate issued by AI to the user.

The CA, acting as AI, validates the certificate request submitted by the BI or by the user. The AI in cooperation with BI will issue the anonymous certificate in the way described next. The AI and BI are communicating using a secure channel. At the end the AI can delegate the issued anonymous certificate (set of certificates) to the user. The certificates will be stored in the OBU of the requested vehicle.

In this model, one of the providers is configured to provide media content such as video and music to registered users from the public who can select and download content on a subscription or per-item basis, typically in accordance with terms of use that may include, for example, digital rights management or other content-protection provisions. In this case,

the certificate authority provides services on a trusted third-party basis for the provider.

## 3. ISSUING AN ANONYMOUS VERIFIABLE PUBLIC KEYCERTIFICATES

The steps of issuing the anonymous verifiable public key certificate are as follows:

**Step 1:** To issue anonymous blind certificate, the user must provide identifying information to the certificate authority. However, the identifying information will typically be handled by the certificate authority to preserve the user's anonymity with the recipient vehicles or by providers. To achieve that, a user authenticates himself to the BI (RA) by, as usual, presenting his identity $ID_v$. BI establishes a record in a database for the user and uses some nonconventional techniques to generate a local identity $ID_{v\text{-}local}$ for the owner. $ID_{v\text{-}local}$ will serve as a (database) key for the record. The $ID_{v\text{-local}}$ value must not be generated in a way that permits an external entity, including AI, to predict the real ID of the user from this value. BI can, for example, calculate the hash of $ID_v$ or the user name and then blinds it by a random number $N_{v\text{-}blind}$. Associated with the $ID_{v\text{-local}}$ in this database is an expiration date. The expiration date is used by BI and AI to reject session-level replay attacks in some exchanges.

BI generates the signal

$$M_V = ID_v \parallel ID_{v-local} \parallel ID_{BI} \parallel Expiration\ date$$

BI and AI use threshold signature scheme, e.g. DAS-based threshold signature scheme to sign $M_v$. At first BI blinds it using a function *Blind*, signs the blinded $M_v$ using his part of the private key, then sends the blinded partially signed message together with the $ID_{v\text{-}local}$ to AI. AI uses the second part of the private key to complete signing the message $M_v$ and sends it back to BI attached with it $ID_{v\text{-local}}$.

BI used the relation:

$$Blind_{inverse}(S(Blind(X)) = S(X),$$

Where $Blind_{inverse}$ is the inverse blind function.

The signed message $[M_v]_{signed}$ is used as "*Token"*. BI sends the *Token* together with $ID_{v\text{-}local}$ to AI. The *Token* is used to verify that a request to the AI has been submitted by a user who is ready registered with BI and has a record at the database of BI contains its real ID. This allows AI, and BI to cooperate to trace the user.

AI uses the *Token* as entry key for the anonymous certificates that he will issue for the user.

If the user needs to communicate after this step directly with AI, BI may supply him with $[M_v]_{signed}$ and the $ID_{v\text{-}local}$. BI may apply directly to AI on behalf of the user to issue the blind certificates by sending the *Token.*

BI stores in its database, the *Token*, $ID_{v\text{-}local}$, the number $N_{v\text{-}blind}$ and $ID_v$.

**Step 2:** The user, or BI on his behalf, prepares a certificate request in a standard format. After that the user the user (or BI) sends the request together with the *Token* to AI via a secure channel.

**Step 3:** Receiving the *Token,* the certificate authority, after verifying the request and the *Token,* starts the blind certificate generation by randomly selecting a subset of key pairs for the user from a fixed-size PKI key set. The particular size of the fixed-size PKI key set can vary by implementation. However, the number of key pairs in the set is selected so that key pairs will be intentionally reused among the users. Such reuse ensures global anonymity for the users by virtue of keyset collisions. The subset of key pairs that the certificate authority assigns to the particular user will similarly be of a fixed-size.

Each certificate in this user certificate set will typically describe the access and usage rights of the user, or other limitations and terms of use/service with respect to a given provider. However, to ensure anonymity for the user, the description will typically not by itself reveal uniquely distinguishable information. The certificates in the set will also include respective public keys from the subset of key pairs.

AI starts by generating for user $i$ a set of *Authenticators* corresponds to the set of keys ($j$ = 1,.., $m$) allocated to him:

$$Authenticator_{ij} = E_{SK_{CA2}}(ID_{v-local} \parallel Token \parallel ID_{BI} \parallel ID_{AI} \parallel PK_{ij} \parallel SK_{ij} \parallel T_i \parallel life \parallel nonce_i)$$

AI uses his own private key to sign the *Authenticator.* This key, also its corresponding public key, is not known for any one. The public key $PK_{CA2}$ is used only by AI when the time comes to link a message to a user. The *nonce* is added as a unique piece of information that only AI knows it.

AI forms $m$ certificates corresponding to the $m$ key pairs allocated to the user. The subject of each certificate is one of the public keys $PK_{ij}$ and has the corresponding $Authenticator_{ij}$ *as part.* For each certificate, the AI then computes a hash over the certificate structure and blinds the hash value. The AI can blind the hash value using his secret private key $SK_{CA2}$. The AI sends the *Token* of the user together with the blinded hash functions for BI for partial signature.

**Step 3:** BI receives the *Token* and the set of blinded certificates hashes via secure channel. After verification, BI uses its share of the threshold private signature key to sign the blind certificates hashes and sends them back to AI.

**Step 4:** Upon receipt of the partially signed certificates hashes, AI matches the *Token* against its list of outstanding requests to the BI. The AI then "un-blind" the received hashes by using his secret public key. This reveals the partially signed certificates hashes. AI uses his part of the split private key to complete the signature of the certificates for the user.

AI records the set of certificates and the *Token* value in its database, to enable later tracing of the certificates to the real user *ID,* if needed.

**Step 5:** The certificate authority provides the user, upon submitting his *Token,* on-line the certificate set along with the private keys from the subset of key pairs. The certificates are stored in the OBU of the user.

**Step 6:** When the user attempts to send a message to other vehicles or applying for service from the provider, his OBU will

randomly select a certificate from the user certificate set to present to the recipients as part of the authentication of the user. In case of VC, the transmitting vehicle will use the private key corresponding to the selected certificate to encrypt the message he wants to send. The recipients (the BOU of other vehicles or the provider) will verify the certificate as being legitimate using conventional techniques and will use the public key to decrypt the sent message.

In the case of sending a safety message to other vehicles, the receiving vehicles may accept the message if they trust the sending vehicle otherwise they can follow some procedures to eliminate the transmitting vehicle.

In case when the user is asking for services from a provider, the provider binds the requested content to the public key from the certificate in order to encrypt it. The encrypted content is transmitted to the client where it is decrypted by the client using the appropriate private key.

Because the PKI key set that the certificate authority uses to generate certificates is of a fixed size, it can be expected that some users will be provided with certificates that have the same public key. Depending on the size of the PKI key set and the number of users that are serviced by the certificate authority, the same public key may be reused in certificates that are issued, for example, to hundreds, thousands, or even a larger number of users. The degree of public key reuse can be varied according to the number of vehicles the registration centre is servicing and the size of the PKI key set. The reuse of public keys among a number of users will make it difficult for anyone, including the case of providers, to uniquely identify or track the user because of the collisions among the commonly-used public keys.

The keyset collisions among multiple users thus provide global anonymity when authenticating with any other vehicle or any tracker. In case of using RA to supply the vehicles with their certificate sets, the possibility of inappropriate or malicious actions that are targeted at a user is almost nil.

## 4. MISBEHAVED VEHICLES/ USERS and THE REVOCATION LIST

In some cases when the user is misbehaved (not trusted by other vehicles or terminated a subscription to the service provider), it will be necessary for the CA to place the user on a revocation list.

The vehicles or the provider must send to the CA a request for revocation attached with it one, or more, of the certificates of the user under consideration. The request must be attached with enough evidences to justify the assertion by the aggrieved party. The AI verifies the assertion by an aggrieved party that a user has abused the anonymity provided by his certificate. After verifications, AI starts to use the certificate (the serial number on the certificate) of revoked user to get his *Token.* Using the *Token* as entry for its database, the CA can define the certificate set issued to the revoked user and the corresponding public keys.

The CA will then add the revoked user's public keys to a revocation list. The revocation list is published to the providers directly and to the other vehicles on the roads. In the last case, CA can communicate with the vehicles through the RSUs.

If the revoked user attempts to access the provider or send safety message to other vehicles using one or more certificates from the user original certificate set, then the recipient will deny access because the user's keys are on the revocation list.

Because of reusing the key by more than one user, the other users that share the key with the revoked user will be also affected; denied by the provider or other vehicles. This situation arises when an unintended revoked user attempts to access content or services from the provider or send a safety message by presenting one certificate from the user's current certificate set. The provider (or the other vehicles) will deny access to the user because the public keys associated with the certificates are included in the revocation list.

However, as the user was not intended to be revoked, the user can simply make a request to the certificate authority for new blind certificates to replace the current user certificate set. The user must send with the request his *Token*. When the certificate authority receives the request, it will check to see if the user is legitimately revoked or not. Because the user is not legitimately revoked in this scenario, the certificate authority will provide a new user certificate set having new public keys along with the corresponding private keys to the user.

The user can start using his new set of certificates to access the provider or to send messages to other vehicles.

If a legitimately revoked user makes a request to receive new certificates in a similar manner as the unintended revoked user, the certificate authority, recognizing his *Token,* will not accept the request as a check will reveal that the user was intended to be revoked.

## 5. MAPPING A CERTIFICATE TO THE REAL ID

The procedure of mapping a certificate to a user's real identity differs from that of revoking. Also the result of the two processes are different; revoking is stopping a user from using the issued set of certificate without identifying his read identity, mapping is targeting getting the real identity of the user.

The mapping process consists of four steps:

**Step 1:** The AI verifies the assertion by an aggrieved party that a user abused the anonymity provided by his blind certificate. The procedure used by AI to verify that can be a court order or any other agreed procedures. If AI found enough evidences, it starts by revoking the certificate set issued to the user. From the serial number on the certificate, AI can find in its database the set of certificate issued to the user and the revokes the certificates. The AI lists the certificates on the Certificate Revocation List.

**Step 2:** The AI searches its database to locate the *Token* that was signed jointly by AI and BI for the user during the issuing process.
The AI passes the *Token* to the aggrieved party using a secure channel.

**Step 3:** The aggrieved party passes the *Token* to BI using secure channel.

**Step 4:** BI verifies his signature on the *Token,* to verify that this is the one generated by it.

BI search its database using the $ID_{v\text{-}local}$ extracted from the token. BI retrieves the user's real ID and provides it to the aggrieved party.

## 6. CONCLUSIONS

In this paper we discussed some of the solutions in use to solve In this paper we discussed the problem of authentication while keeping privacy. We discussed some of the existing solutions. We used in this paper to concepts to achieve improvement on the current solutions: sue of split Trusted Authority and allowing issuing the same key for more than on user, i.e. more than one certificate has the same entity subject. We showed that tends to improve the security and the privacy.

## 7. REFERENCES


[1] P. Papadimitratos, L. Buttyan, T. Holczer, E. Schoch, J. Freudiger, M. Raya, Z. Ma, F. Kargl, A. Kung and J.-P. Hubaux, "Secure vehicular communication systems: design and architecture - [topics in automotive networking]", **IEEE Communications Magazine**, vol.46, no.11, November 2008.

[2] F. Kargl, P. Papadimitratos, L. Buttyan, M. Muter et al., "Secure Vehicular Communication Systems: Implementation, Performance, and Research Challenges - [topics in automotive networking]", **IEEE Communications Magazine**, vol.46, no.11, November 2008.

[3] L. Buttyan, T. Holczer, and I. Vajda, "On the Effectiveness of Changing Pseudonyms to Provide Location Privacy in VANETs," **Proc. ESAS, F. Stajano, et al., Eds.**, vol. 4572, Springer Verlag, July 2007, pp. 129–41.

[4] J.-P. Hubaux, S. Capkun, and J. Luo, "The Security and Privacy of Smart Vehicles," **IEEE Security & Privacy Magazine**, vol. 2, no. 3, pp. 49–55, 2004.

[5] M. Myers, R. Ankney, A.Malpani, et al, "**X.509 Internet Public Key Infrastructure On-line Certificate Status Protocol-OCSP**", RFC 2560, June 1999.

[6] P. MacKenzie, K. Reiter, "Two-Party Generation of DSA Signature", **Crypto 2001**.

[7] Shaohua Tang, "Simple Threshold RSA Signature Scheme Based on Simple Secret Sharing," in "**Computational Intelligence and Security**," CIS 2005, Part II, Springer, pp.186-191, 2005

# An Efficient Authenticated Key Agreement Scheme Without Using Smart Card

Chun-Ta Li
Department of Information Management
Tainan University of Technology
529 Jhongjhen Road, Yongkang, Tainan, Taiwan 710, R.O.C.

Min-Shiang Hwang
Department of Computer Science & Information Engineering
Asia University
500, Lioufeng Rd., Wufeng, Taichung 41354, Taiwan, R.O.C.
Email: mshwang@asia.edu.tw

Pin-Chieh Huang
Department of Electrophysics[§]
National Chiao Tung University
1001 University Road, Hsinchu, Taiwan 300, R.O.C.

## Abstract

*Authenticated key agreement based on passwords over insecure networks is the conventional method of secure communications in the various networking environments. In this article, we propose an efficient authenticated key agreement scheme without using smart card and the security of our proposed scheme is based on exclusive OR operation, hashing function, and discrete logarithm problem. Therefore, the proposed scheme does not need the use of any additional public-key infrastructure and it is not only secure against security attacks but also is more efficient than the other schemes.*

**Keyword:** *Authenticated Key Agreement, Cryptography, Modification Attack, Network Security.*

## 1. Introduction

In 1976, Diffie and Hellman [1] first proposed a well-known key agreement scheme that the two communication parties can agree a common session key in an insecure network [2, 12, 13, 15, 16]. However, the Diffie-Hellman scheme suffers from the man-in-the-middle attack [4, 8, 18, 19]. In 1999, Seo and Sweeney [22] proposed an authenticated key agreement scheme (SAKA) in which two communication parties used a pre-shared password to achieve user authentication, but Tseng [23] demonstrated that it could not withstand the replay attack in which a malicious user could cheat an honest party into believing a wrong common session key. So then, Tseng proposed an improved scheme to against the replay attack.

Later, Ku and Wang [7] indicated that the Tseng scheme suffers from the backward replay attack without modification and modification attack [11, 14, 17]. Then, Ku and Wang proposed an improved scheme to dispose of these two attacks. In 2003, Hsu et al. [3] pointed out that the Ku-Wang scheme is vulnerable to the modification attack and also proposed an improved scheme that not only enhances the security of the Ku-Wang scheme but also is more efficient than the previous schemes. However, in 2004, Lee and Lee [10] showed that the Hsu et al. scheme is still insecure to the modification attack and further proposed an improved scheme. Lee-Lee's scheme repaired the weakness of Hsu et al. scheme and it is as efficient as the Hsu et al. scheme. Unfortunately, in 2005, Kim et al. [6] showed that the Lee-Lee scheme was breakable by guessing attack. Moreover, Lee et al. [9] also presented that the Lee-Lee scheme was breakable by man-in-the-middle attack. In this article, we proposed an improved key agreement scheme without using smart card and the improved scheme is not only suggested to eliminate the weaknesses

**Table 1. Notations used in this article**

| | |
|---|---|
| Alice, Bob | Two communication parties |
| $id_A, id_B$ | Identities of Alice and Bob |
| $P$ | A large prime number |
| $PW$ | A common password shared between Alice and Bob |
| $Q$ | An integer pre-computed from $PW$ |
| $Q^{-1}$ | The inverse of $Q \pmod P$ |
| $g$ | A base generator $\in Z_P^*$ with the order $P-1$ |
| $a$ | A random number chosen by Alice |
| $b$ | A random number chosen by Bob |
| $K$ | A common session key derive from Alice and Bob |
| $H(\cdot)$ | One-way hash function |

in Lee-Lee scheme but also is more efficient than previously proposed schemes [5, 9, 21, 20] in terms of computation and communication loads.

The article is organized as follows. First, we propose a simple improved scheme in Section 2. In Section 3, we analyzed the security of our improved scheme and compared it with other related schemes. Finally, we conclude this article in Section 4.

## 2. The Proposed Scheme

In this section, we will propose a simpler key agreement scheme based on the Lee-Lee scheme [10]. The notations in Table 1 are used in this article.

The detailed steps of the proposed scheme are described as follows and in Figure 1.

**Step 1:** Alice computes $X_1 = g^a \oplus Q \bmod P$ and sends $X_1$ to Bob.

**Step 2:** After receiving the message $X_1$, Bob computes $Y_1 = g^b \oplus Q \bmod P$.

**Step 3:** Next, Bob computes the session key as follows:

$$X = X_1 \oplus Q \bmod P = g^a \bmod P$$
$$K_2 = X^b \bmod P = g^{ab} \bmod P.$$

**Step 4:** Lastly, Bob checks whether $K_2 \neq 1$ holds or not. If it holds, Bob computes $Y_2 = H(id_b, X_1, K_2)$ and sends $Y_1$ and $Y_2$ to Alice. Otherwise, the key agreement scheme is terminated.

**Step 5:** After receiving the messages, $Y_1$ and $Y_2$, Alice first computes the session key as follows:

$$Y = Y_1 \oplus Q \bmod P = g^b \bmod P.$$
$$K_1 = Y^a \bmod P = g^{ab} \bmod P.$$

**Step 6:** Then, Alice verifies $Y_2 \stackrel{?}{=} H(id_b, X_1, K_1)$ and checks whether $K_1 \neq 1$ holds or not. If above holds, Alice computes $X_2 = H(id_a, Y_1, K_1)$ and sends $X_2$ to Bob. If it does not hold, it means that Alice and Bob can not agree a common session key and the key agreement scheme is terminated.

**Step 7:** After receiving the message $X_2$, Bob verifies $X_2 \stackrel{?}{=} H(id_a, Y_1, K_2)$. If it holds, Alice and Bob are now confirmed that the common session key $K = K_1 = K_2 = g^{ab} \bmod P$. Otherwise, the key agreement scheme is terminated as previous circumstance shows.

## 3. Analysis

In this section, we analyze the security of the proposed scheme and compare the related works with ours in terms of computation and communication loads as follows.

### 3.1 Security Analysis

1. In our scheme, it is difficult for an attacker to derive the pre-computed integer $Q$ from receives messages $X_1, Y_1, Y_2$ and $X_2$ because the complexity of computing $Q$ from receives messages is a discrete logarithm problem. Therefore, our scheme is secure to against guessing attack.

2. An attacker may try to alter the messages to cheat both of communication parties into believing a wrong session key(modification attack). It does not work unless he/she knows the common session key $K$.

3. If an attack Eve tries to masquerade Alice and cheat Bob(masquerade attack), Eve can send the deceitful message $X_1' = g^{a'} \oplus Q'$ to Bob, where $a'$ is randomly selected and $Q'$ is derived from a guessed password $P'$. Then, Eve wants to verify whether her guess holds or not, she must check $K_2 = (g^{a'} \oplus Q' \oplus Q)^a \bmod P)$. However, Eve only has $(g^{a'} \oplus Q')$ and without knowing $a$. As a result, it is difficult for Eve to masquerade Alice and cheat Bob because she has to solve the discrete logarithm problem.

### 3.2 Performance Analysis

In this section, we compare the efficiency of the proposed scheme, the Lu-Cao scheme [20], the Ryu et al. scheme [21], the Lee et al. scheme [9], and the Hwang et al. scheme [5]. Previous schemes are briefly reviewed below.

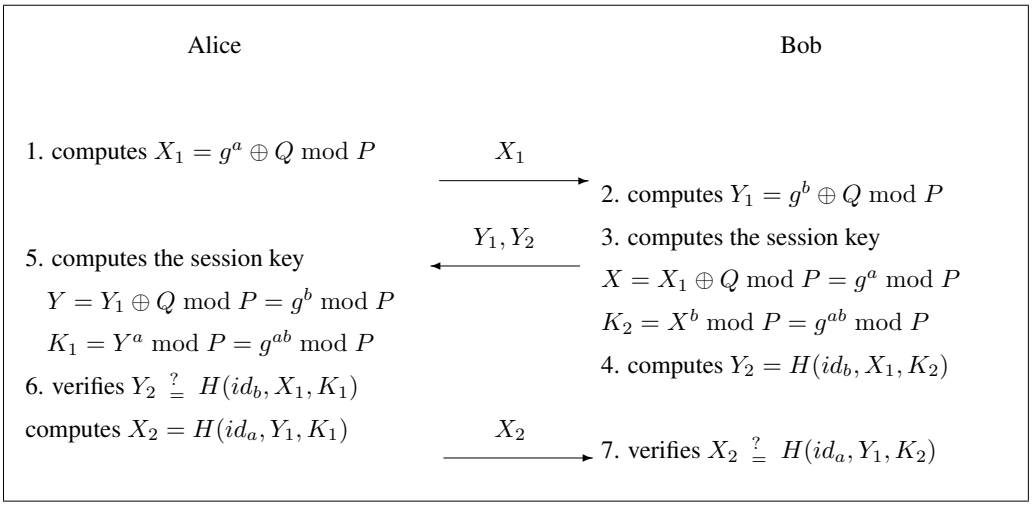



**Figure 1. The proposed scheme**

- The Hwang et al. Scheme [5]: There are two steps in the Hwang et al. scheme. First, the two communication parties, Alice and Bob already generated their long-term secret key $X_a$, $X_b$ and generated two short-term secret keys $\{R_{a1}, R_{a2}\}$ and $\{R_{b1}, R_{b2}\}$, respectively. After two steps messages transmitted, the two communication parties, Alice and Bob are now confirmed as the four common session keys by the following equations:

$$\begin{aligned} K_{1a} &= K_{1b} = g^{Ra1Rb1} \bmod P, \\ K_{2a} &= K_{2b} = g^{Rb1Ra2} \bmod P, \\ K_{3a} &= K_{3b} = g^{Ra1Rb2} \bmod P, \\ K_{4a} &= K_{4b} = g^{Ra2Rb2} \bmod P. \end{aligned}$$

- The Ryu et al. Scheme [21]: Before the scheme begins, Alice and Bob pre-shared a password $Q$ and knew the system parameters, including a large prime $P$ and its generator $g$. First, Alice first chooses a random number $a$, and computes $X = g^a + Q \bmod P$ and then sends $X$ to Bob. After receiving the message $X$, Bob chooses a random number $b$, and computes $Y = g^b \bmod P$, $Y_1 = (X - Q)^b \bmod P$, and $Y_2 = H(ID_a, X, Y_2)$. Then Bob sends $Y$ and $Y_2$ to Alice. After receiving the message $Y$, $Y_2$, Alice computes $X_1 = Y^a \bmod P$ and checks if $H(ID_a, X, X_1)$ equals to $Y_2$ or not. If it holds, Alice computes $X_2 = H(ID_b, Y, X_1)$ and the common session key $K_1 = kdf(ID_a, ID_b, X_1)$ and sends $X_2$ to Bob, where $kdf(.)$ is a key derivation function. Similarly, Bob could verify the validity of $X_2$. If $X_2$ equlas to $H(ID_b, Y, Y_2)$, Bob computes the common session key $K = K_1 = K_2 = kdf(ID_a, ID_b, Y_1)$.

- The Lee et al. Scheme [9]: There are two phases in the Lee-Lee scheme, *Key establishment phase* and *Key validation phase*, respectively. Before the scheme begins, Alice and Bob publish $g^X$ and $g^Y$, respectively(where $X = aQ$ and $Y = bQ$). In *Key establishment phase*, Alice first chooses a random number $a$, and computes $X = aQ$, $X_1 = g^X \bmod P$ and then sends $X_1$ to Bob. After receiving the message $X_1$, Bob chooses a random number $b$, and computes $Y = bQ$, $Y_1 = g^Y modP$ and then sends $Y_1$ to Alice. After receiving the message $Y_1$, Alice computes the common session key $K_1 = Y_1^{Q^{-1}a}$. Similarly, Bob could compute the common session key $K_2 = X_1^{Q^{-1}b}$. After that, the two communication parties, Alice and Bob could derive the common session key $K = K_1 = K_2 = g^{ab} \bmod P$. In *Key validation phase*, in order to convince the validity of the derived session key, Alice and Bob should reciprocally carry out the follow-

ing steps. First, Alice checks whether $K_1 \neq 1$ holds or not. If it holds, Alice computes $X_2 = H(id_A, X_1, K_1)$ and sends $X_2$ to Bob. Then, Bob verifies the validation of the equation $X_2 \stackrel{?}{=} H(id_A, X_1, K_2)$. If it holds, Bob checks whether $K_2 \neq 1$ holds or not. If it holds, Bob computes $Y_2 = H(id_B, Y_1, K_2)$ and sends $Y_2$ to Alice. Finally, Alice could verify the validation of the equation $Y_2 \stackrel{?}{=} H(id_B, Y_1, K_1)$. If it holds, the two communication parties, Alice and Bob are now confirmed that the common session key $K = K_1 = K_2 = g^{ab} \bmod P$.

- The Lu-Cao Scheme [20]: There are three communication steps in Lu-Cao scheme. In first step, Alice chooses a random number $a$, and computes $X_1 = g^{aQ} \bmod P$ and $X_2 = g_1^a \bmod P$, then sends messages $X_1$, $X_2$ to Bob. On receiving $X_1$ and $X_2$, Bob first computes $X_1' = X_1^{Q^{-1}} = g^a \bmod P$, and then chooses two random numbers $b_1$ and $b_2$ to computes $Y_1 = g^{b_1} g_1^{b_2} \bmod P$ and $Y_2 = X_1'^{b_1} X_2^{b_2} \bmod P$. In second step, Bob sends $Y_1$ and $Y_3 = H(A||B||X_1||X_2||Y_1||Y_2||0)$ to Alice. After receiving Bob's message, Alice first computes $Y_2' = Y_1^a \bmod P$ and then verifies $Y_3 \stackrel{?}{=} H(A||B||X_1||X_2||Y_1||Y_2'||0)$. If it holds, Alice authenticates Bob. Otherwise, the protocol is terminated. In third step, Alice sends $X_3 = H(A||B||X_1||X_2||Y_1||Y_2'||1)$ to Bob and computes the common session key $K_1 = H(A||B||X_1||X_2||Y_1||Y_2')$. On receiving $X_3$, Bob verifies $X_3 \stackrel{?}{=} H(A||B||X_1||X_2||Y_1||Y_2||1)$. If it does hold, Bob authenticates Alice. Finally, Bob computes the common session key $K_2 = H(A||B||X_1||X_2||Y_1||Y_2)$.

Next, we compare the efficiency of our scheme and previous related schemes in Table 2. As Table 2 shows, in terms of the computation loads, our proposed scheme is more efficient than other schemes that only four exponential operations are required and the computation loads of Hwang et al. scheme is the highest because eighteen exponential operations are required.

Furthermore, from the perspective of communication load shows, the Hwang et al. scheme is eight messages sent, two communication steps and four random numbers; The Lu-Cao scheme is five messages sent, three communication steps and three random numbers; The communication cost of the Lee et al. scheme is six messages sent(two publish messages are included), four communication steps and two random numbers; The communication cost of our scheme is four messages sent, three communication steps and two random numbers. Similarly, in terms of communication loads, our proposed scheme is more efficient than the other related schemes mentioned above.

## 4. Conclusion

In this article, we have proposed an efficient authenticated key agreement scheme. We also give a comparison with our scheme and some related schemes in terms of communication and computation loads. From the performance result shows, the communication and computation loads of the proposed scheme are the lowest and it is as secure as the other related schemes mentioned above.

## 5. Acknowledgments

We wish to thank many anonymous referees for their suggestions to improve this paper. This work was supported in part by National Science Council under the grants NSC 98-2221-E-005 -050 -MY3.

## References

[1] Whitfield Diffie and M. Hellman. New directions in cryptology. *IEEE Transations on Information Theory*, IT-22(6):644–654, 1976.

[2] Marko Holbl and TITLE = Tatjana Welzer.

[3] Chien-Lung Hsu, Tzong-Sun Wu, Tzong-Chen Wu, and Chris Mitchell. Improvement of modified authenticated key agreement protocol. *Applied Mathematics and Computation*, 142(2-3):305–308, 2003.

[4] Min-Shiang Hwang, Chih-Wei Lin, and Cheng-Chi Lee. Improved yen-joye's authenticated multiple-key agreement protocol. *IEE Electronics Letters*, 38(23):1429–1431, 2002.

[5] Ren-Junn Hwang, Sheng-Hua Shiau, and Chih-Hua Lai. An enhanced authentication key exchange protocol. In *Proceedings of the 17th International Conference on Advanced Information Networking and Applications (AINA'03), IEEE*, pages 202–205, 2003.

[6] Kee-Won Kim, Eun-Kyung Ryu, and Kee-Young Yoo. Cryptanalysis of lee-lee authenticated key agreement scheme. *Applied Mathematics and Computation*, 163(1):193–198, 2005.

[7] Wei-Chi Ku and Sheng-De Wang. Cryptanalysis of modified authenticated key agreement protocol. *IEE Electronics Letters*, 36(21):1770–1771, 2000.

[8] Cheng-Chi Lee, Min-Shiang Hwang, and Li-Hua Li. A new key authentication scheme based on discrete logarithms. *Applied Mathematics and Computation*, 139(2):343–349, 2003.

**Table 2. Performance comparison**

| | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 |
|---|---|---|---|---|---|---|---|---|---|
| Hwang et al. [5] | 8 | 2 | 4 | 2 | 18 | 4 | 0 | 4 | 4 |
| Ryu et al. [21] | 4 | 3 | 2 | 0 | 4 | 0 | 6 | 0 | 1 |
| Lee et al. [9] | 6 | 4 | 2 | 0 | 6 | 2 | 4 | 0 | 1 |
| Lu-Cao [20] | 5 | 3 | 3 | 0 | 8 | 1 | 6 | 0 | 1 |
| The proposed scheme | 4 | 3 | 2 | 0 | 4 | 0 | 4 | 4 | 1 |

*Notes.* C1: needs to send messages; C2: communication steps; C3: number of random numbers; C4: number of addition operation; C5: number of exponential operation; C6: number of multiplication operation; C7: number of $H(\cdot)$ operation; C8: number of XOR operation; C9: number of common session key.


[9] Keon-Jik Lee and Byeong-Jik Lee. Cryptanalysis of the modified authenticated key agreement scheme. *Applied Mathematics and Computation*, 170(1):280–284, 2005.

[10] N. Y. Lee and M. F. Lee. Further improvement on the modified authenticated key agreement scheme. *Applied Mathematics and Computation*, 157(3):729–733, 2004.

[11] Chun-Ta Li and Yen-Ping Chu. Cryptanalysis of threshold password authentication against guessing attacks in ad hoc networks. *International Journal of Network Security*, 8(2):166–168, 2009.

[12] Chun-Ta Li and Min-Shiang Hwang. An efficient biometrics-based remote user authentication scheme using smart cards. *Journal of Network and Computer Applications*, 33(1):1–5, 2010.

[13] Chun-Ta Li and Min-Shiang Hwang. An online biometrics-based secret sharing scheme for multiparty cryptosystem using smart cards. *International Journal of Innovative Computing, Information and Control*, accepted, 2009.

[14] Chun-Ta Li, Min-Shiang Hwang, and Yen-Ping Chu. Improving the security of a secure anonymous routing protocol with authenticated key exchange for ad hoc networks. *International Journal of Computer Systems Science and Engineering*, 23(3):227–234, 2008.

[15] Chun-Ta Li, Min-Shiang Hwang, and Yen-Ping Chu. A secure and efficient communication scheme with authenticated key establishment and privacy preserving for vehicular ad hoc networks. *Computer Communications*, 31(12):2803–2814, 2008.

[16] Chun-Ta Li, Min-Shiang Hwang, and Yen-Ping Chu. An efficient sensor-to-sensor authenticated path-key establishment scheme for secure communications in wireless sensor networks. *International Journal of Innovative Computing, Information and Control*, 5(8):2107–2124, 2009.

[17] Chun-Ta Li, C. H. Wei, and Y. H. Chin. A secure event update protocol for peer-to-peer massively multiplayer online games against masquerade attacks. *International Journal of Innovative Computing, Information and Control*, accepted, 2009.

[18] Eric Jui-Lin Lu and Min-Shiang Hwang. An improvement of a simple authenticated key agreement algorithm. *Pakistan Journal of Applied Sciences*, 2(1):64–65, 2002.

[19] Eric Jui-Lin Lu, Cheng-Chi Lee, and Min-Shiang Hwang. Cryptanalysis of some authenticated key agreement protocols. *International Journal of Computational and Numerical Analysis and Applications*, 3(2):151–157, 2003.

[20] Rongxing Lu and Zhenfu Cao. Off-line password guessing attack on an efficient key agreement protocol for secure authentication. *International Journal of Network Security*, 3(1):35–38, 2006.

[21] E. K. Ryu, K. W. Kim, and K. Y. Yoo. An authenticated key agreement protocol resistant to a dictionary attack. In *Proceedings of ICCSA 2004*, volume LNCS 3043, pages 603–610, 2004.

[22] D. Seo and P. Sweeney. Simple authenticated key agreement algorithm. *IEE Electronics Letters*, 35(13):1073–1074, 1999.

[23] Yuh-Min Tseng. Weakness in simple authenticated key agreement protocol. *IEE Electronics Letters*, 36(1):48–49, 2000.

# Rigorous Calculation of the Partition Function for the Finite Number of Ising Spins

**Alexey A. Peretyatko, Ivan A. Bogatyrev, Yury V. Kirienko,**
**Konstantin V. Nefedev, Valery I. Belokon**
**Far Eastern Federal University**
**School of Natural Science**
**Russia,Vladivostok, Sukhanova str.8-43**
knefedev@phys.dvgu.ru

## ABSTRACT

The high-performance scalable parallel algorithm for rigorous calculation of partition function of lattice systems with finite number Ising spins was developed. The parallel calculations run by C++ code with using of Message Passing Interface and massive parallel instructions. The algorithm can be used for the research of the interacting spin systems in the Ising models of 2D and 3D. The processing power and scalability is analyzed for different parallel and distributed systems. Different methods of the speed up measuring allow obtain the super-linear speeding up for the small number of processes. Program code could be useful also for research by exact method of different Ising spin systems, e.g. system with competition interactions.

**Keywords:** Massive-Parallel Algorithm, MPI, C++, Parallel and Disributed Calculations, Ising Model, Partition Function, Statistical Mechanics.

## 1. INTRODUCTION

Among the huge variety of fundamental scientific tasks there are problems from the solution of which ones the comprehension progress of the Nature depends in whole. The solution of Ising model, a simple mathematical model of phase transitions of correlated systems, could be such a problem. The rigorous solution in this model it is necessary not only for the development of a statistical physics, but also and in related areas: nanotechnology, nanomaterial science, nanobiotechnology and et al. Today it is possible to work within systems which ones consist of relative small (finite) number of atoms. The exact solution in Ising model would be very desirable for further development of science and as consequence of a technology. The existing today single-threaded searching methods have strong limitations and not allow obtain the solution within a reasonable period of time. The total amount of a terms partition function grow up as power function with the increasing of the number of elements in the system, moreover Ising model could not be solved exactly by analytical approach in 3D, since the non-planar tasks are NP-complete problem [1-2]. Thus for exact construction of the partition function even for small finite systems it is desirable have a scalable algorithm for the using of computer power which one could support the intensive parallel calculations. The productivity of computers enhances every ten years approximately in thousand times, that essentially expands the class of solvable numerical tasks.

Approximate research methods of 2D and 3D Ising model, such for example as Metropolis algorithm and method Monte Carlo (MC) [3-4] have wide acknowledgement. With respect to rigorous solution the situation is more problematical. The rigorous analytical solution for 1D case was obtained by Ising in 1925, and exact solution for 2D case was obtained by Onsager in 1944 [5] (difference between “exact” and “unrigorous” defined in [6]). In work [7] by authors it was shown that the critical temperatures of transition of finite systems to ferromagnetism could be calculated more precisely than in approximate analytical theories. Therefore it is interesting to construct the rigorous solution for finite number of spins for a large lattices with the number of nearest neighbors Z=4 or more, and make the comparison of results, which ones can be obtained by means of numerical simulation and method rigorous calculation.

The aim of this work is development, optimization and testing on a different computational resource of high-performance scalable parallel algorithm of rigorous calculation of partition function for the finite number $N$ of Ising spins with ferromagnetic exchange on the planar lattice with Z=4, and show the possibility of generalization of this algorithm on 3D arbitrary lattices.

This work was done at support of the Russian Federal Aim Program “Scientific and scientific-pedagogical staff of innovation Russia”, Ministry of Education and Science of Russian Federation, State Contract № 14.740.11.0289, "Development of computer programs based on algorithms of parallel programming and optimization of high-performance distributed computing environment for solving problems of natural science."

## 2. ISING MODEL

Let us shortly remind the main idea of the simplest model of ferromagnetism - the Ising model, which one has not rigorous solution yet in spite of all its simplicity. In this model it permits only the interactions between nearest neighbors usually. The Hamiltonian of the system in general case is

$$H = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{Z} J_{ij} S_i S_j - h\sum_{i}^{N} S_i, \qquad (1)$$

where $h$ is external magnetic field, the summation over $j$ denotes the summation over neighbors (nearest here). Using summands in the Eq. (1) it is possible to put for the each configuration of $2^N$ possible configurations in correspondence two numbers – the energy $E_i$ and the spin excess $M_i$ [8]. Exchange constant $J_{ij}$ is positive and in considered case equals to one (in principle could be sign-changing and distance-

sensitive in general case). The partition function of finite number of Ising spins

$$Z_N(h,T) = \sum_{S_1}\sum_{S_2}\dots\sum_{S_N} Exp\left[-\frac{H}{T}\right]. \quad (2)$$

The troubles with the rigorous numerical calculation of Eq. (2) are in the estimation of the degeneracy multiplicity of the spin excess $M_i$ (total sum of all $S_i$) over energy $E_i$ (total sum exchange energy of all pair spins). The degeneracy means, that to one value of $M_i$ it could correspond a different values of $E_i$. The degeneracy multiplicity equals to the number of combinations with given quantity of pairs of side by side "units" ("zeros") with taking into account the periodical boundary conditions. In other words in the combinatorial terms we need estimate the number of permutations with repeating and limitations on the distribution of side by side elements. The total amount of the degeneracy multiplicities for the fixed $M_i$ over all permitted $E_i$ is the binomial coefficient.

## 3. SCALABLE PARALLEL ALGORITHM

The simplicity of proposed approach gives the possibilities of extension of the elaborated algorithm for 2D Ising lattice, also and on 3D lattice with given number of a nearest neighbors *Z*. The limitation of linear search parallel algorithm is in linear dependence from computer performance.

In general case the lattice can be presented by the set of bit-vectors. For 2D this set bit-vectors $\{a_i\}$ could be the one dimension array, where two nearest elements correspond in two nearest columns of lattice, for 3D it is two dimensional matrix $\{a_{ij}\}$. De facto the information about distribution of spin states each column of the lattice was written it the short unsigned integer variables, number of which ones equals to linear dimension of lattice. Integer variable size is four bytes, i.e. 32 bits. For calculations of a small lattice it needs small number of these bits, hereinafter referred to as significant. A unary and pair logical as well as arithmetical operations with a parts of such bit-vectors, the number of significant bits which ones equal to second linear dimension of lattice, allow obtain the $M_k$ and $E_k$. Total sum of $M_k$ and $E_k$ of all variables allow obtains the $M_i$ and $E_i$ for one configuration. All configurations can be "counted" by parallel changing of value of each integer variable in a diapason from 0 to 2 in power *n*, where *n* is a linear dimension of the lattice.

### a. Spin Excess

The sum of all units of all bit-vectors gives the number of spins, which ones have direction "up". It is easy to define the spin excess as the deference between the total numbers of spins and the double amount of the spin "up". The quantity of units was defined with using of Kernighan algorithm [9]

```
int q_units(unsigned short int i)
{ int j = 0;
 while (i)
 {i&=i-1; ++j;}
 return j; }
```

Really, in code of program it has used the optimization which one connects with the storage in an unsigned short integer cache array $\{b_i\}$ of the quantities units (weights of bit-word) calculated once for the each decimal numbers and reference to the elements of the array in cycles. The reference on the elements of the given array is used instead of an all arithmetical operations, where the enumeration of the bit-vector units takes place.

### b. Rows Exchange Energy

The calculation of exchange energy between spins in rows can be done in three steps:

*1) The operation XOR between all pairs $a_i$ and $a_{i+1}$ , including periodical boundary conditions $a_1 \wedge a_N$, for 2D lattice ($a_{ij} \wedge a_{i+1,j}$ or $a_{ij} \wedge a_{i,j+1}$ for 3D case);*

*2) The invert conversion of the previous operation result, i.e. implimentation of NOT logical operation;*

*3) The reset of bits which ones have not information about spin state of the system. The subtraction from resultant bit-vector $2^{32}$-$2^n$ (for expample order 32 is for "int" type of the bit-word size);*

The quantity of units in the resultant bit-word gives us the number of the pair exchange interactions with positive sign in rows, and should be stored in the separate variable *En*.

### c. Columns Exchange Energy

The enumeration of the positive pair exchange spins interactions in the each $a_i$ ($a_{ij}$) column achieve by the following way:

*1) To each decimal number "j" - index of the integer aray $\{c_j\}$ - put in correspondence the dicimal number $c_j$, the binary notation of which one is circular shift of significant bits of number "j";*

*2) The implemention of the operation XOR between all pairs of integer number $a_i$ and $c_j$, where $j= a_i$;*

*3) The invert conversion (NOT operation) of the previous operation result;*

*4) The reset of bits which ones have not information about spin state of the system;*

*5) The references on elements of the array $\{b_i\}$ is used for the determination of units quantity and summation with En;*

The final value of *En* is the total amount of the all pair exchange interactions with positive sign for one configuration. There are not troubles to calculate the total exchange energy of the spin pairs interactions by using of the simple arithmetical expression $E_{tot}=2(-En+N)$.

### d. Parallelization. MPI

The Massage Passing Interface (MPI) standard extends the possibility of computers and makes available parallel algorithmization that allows the solving of the tasks, the solution of which ones couldn't be obtained in the reasonable time early. The possibility of the parallel execution in our C++ code is implemented via the MPI libraries.

The number the nested loops equals to the length of the vector $\{a_i\}$, i.e. number bit-vectors. The task of the parallelization of the nested iterations requires the application of the technology of dynamical processes generation. In the current version of the program we did not used the dynamical parallelization, but there are some possibilities of the organization of a dynamical generation of the parallel processes. For example, the group communicators can be used.

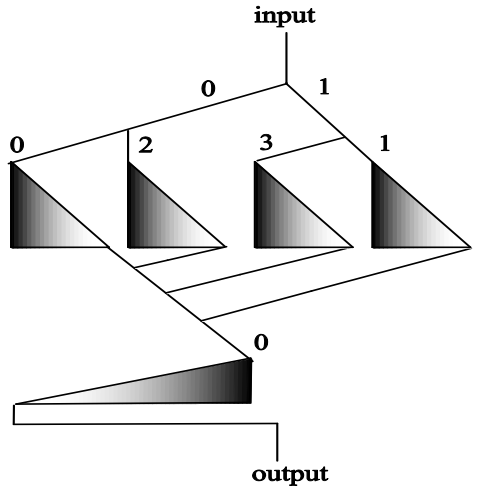



Figure 1. The logical scheme (graph) of the parallel algorithm with dynamical generation of two processes by two.

The communication process topology is on the Fig.1. The given approach is acceptable for the two nested loops

```
for(j=0; j<2;j++)
for(i=0; i<2;i++)
```

where the two processes were generated by two other. The data is distributed uniformly between parent processes, "0" and "1", and the child processes, "2" and "3". The "0" process is chosen as the root, and the calculation results are passed in it after completing of the computing. The root process receives, sorts and saves the resultant data.

## 4. TESTING THE CODE PERFORMANCE FOR DIFFERENT COMPUTER SYSTEMS

To find the fastest solution of the resource-intensive tasks supercomputers with multi-core architecture are used. However, such the machines are quite expensive and quickly become outdated without the possibility of upgrading. Usually the architecture is push-typed and has weak capability to modernization. It is possible to create cloud computing power with using of the regular workstations with heterogeneous architectures. Therefore, a heterogeneous, scalable computing system for the testing of the performance our algorithm was used. Also the work of program was checked on HP Cluster (16 and 32 cores).

### *GNU / Linux PelicanHP*

We used two computers: Console Cluster - Intel Core 2 Quad 2,5 Ghz, 4Gb DDR2, 500Gb HDD, 9800GTX 512Mb, 10/100/100 Mb Ethernet; cluster nodes: AMD Athlon 64 X2 2,8 GHz, 3 Gb DDR2, 232 Gb HDD, ATI Radeon HD 4850, 10/100/100 Mb Ethernet. The used GRID is the distribution GNU / Linux PelicanHPC x86_64 (v2.3.2) linux kernel 2.6.32-5, [10]; implementation MPI: OpenMPI v1.4.2-4. The loading of the operating system made from USB drive on the console of control computer. Service DHCP runs for a dynamic IP-addressing of the cluster nodes.

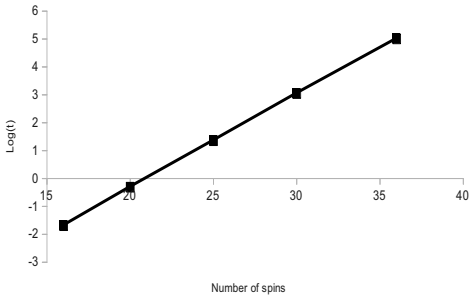



Figure 2. The dependence on logarithm of time (sec) from number of the spins.

The service runs NFS (Network File System) to create a shared resource across the network and to boot image is installed on Linux NFS, that one makes possible to run the cluster nodes over the network (PXE boot). After of the searching for available nodes the cluster is installed.

We run our program on 2 cores which ones is the ownership of two different physical processors at different motherboards connected directly to RJ-45 cable. On the Fig. 2 are the results of estimates of the computation time in dependence on the number of spins. According to the calculations it is possible to make prediction about the time of calculation of the 7x7 system. On 2 cores the time for system of 49 spins is about 76 years! It is easy to see that the scaling of the algorithm on more number of cores can give the solution in reasonable time.

### *HP Cluster*

To check the code performance it is used of HP cluster Intel Xeon E5410 @ 2.33GHz. The 16 cores have used for estimation of the executing time dependence from number of cores. There are time logarithms of program running (in sec.) on 16 cores for different numbers of spins (36 and 30 spins) on the Fig. 3. Unfortunately the available resources couldn't give us the possibilities to estimate of the solution time for bigger number of spins.

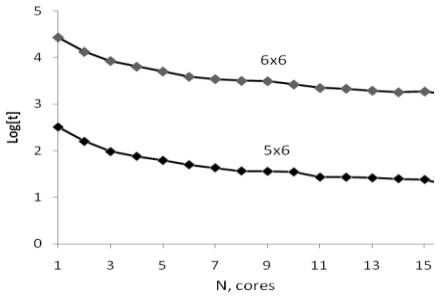



Figure 3. The dependence on logarithm of time from number of the cores for defferent number of spins.

The elaborated algorithm for the rigorous calculation of partition function for finite number of Ising spins on the square lattice displays the scalability. Processes are independent and run approximately the same time.

TABLE 1. Results of calculation for 4x4 system.

| *Nc* | $M_i$ | $E_i$ | *Nc* | $M_i$ | $E_i$ | *Nc* | $M_i$ | $E_i$ |
|---|---|---|---|---|---|---|---|---|
| | | | 64 | 2 | -12 | | | |
| 1 | 16 | -32 | 624 | 2 | -8 | 96 | -4 | -12 |
| 16 | 14 | -24 | 1920 | 2 | -4 | 704 | -4 | -8 |
| 32 | 12 | -20 | 3680 | 2 | 0 | 1824 | -4 | -4 |
| 88 | 12 | -16 | 3136 | 2 | 4 | 2928 | -4 | 0 |
| 96 | 10 | -16 | 512 | 2 | 12 | 1568 | -4 | 4 |
| 256 | 10 | -12 | 1392 | 2 | 8 | 768 | -4 | 8 |
| 208 | 10 | -8 | 96 | 2 | 16 | 64 | -4 | 12 |
| 24 | 8 | -16 | 16 | 2 | 24 | 56 | -4 | 16 |
| 256 | 8 | -12 | 8 | 0 | -16 | 192 | -6 | -12 |
| 736 | 8 | -8 | 768 | 0 | -8 | 688 | -6 | -8 |
| 228 | 8 | 0 | 1600 | 0 | -4 | 1664 | -6 | -4 |
| 576 | 8 | -4 | 4356 | 0 | 0 | 1248 | -6 | 0 |
| 192 | 6 | -12 | 2112 | 0 | 8 | 448 | -6 | 4 |
| 688 | 6 | -8 | 3264 | 0 | 4 | 128 | -6 | 8 |
| 1664 | 6 | -4 | 576 | 0 | 12 | 24 | -8 | -16 |
| 1248 | 6 | 0 | 120 | 0 | 16 | 256 | -8 | -12 |
| 448 | 6 | 4 | 64 | 0 | 20 | 736 | -8 | -8 |
| 128 | 6 | 8 | 2 | 0 | 32 | 576 | -8 | -4 |
| 96 | 4 | -12 | 64 | -2 | -12 | 228 | -8 | 0 |
| 704 | 4 | -8 | 624 | -2 | -8 | 96 | -10 | -16 |
| 1824 | 4 | -4 | 1920 | -2 | -4 | 256 | -10 | -12 |
| 2928 | 4 | 0 | 3680 | -2 | 0 | 208 | -10 | -8 |
| 1568 | 4 | 4 | 3136 | -2 | 4 | 32 | -12 | -20 |
| 768 | 4 | 8 | 1392 | -2 | 8 | 88 | -12 | -16 |
| 56 | 4 | 16 | 512 | -2 | 12 | 16 | -14 | -24 |
| 64 | 4 | 12 | 96 | -2 | 16 | 1 | -16 | -32 |
| | | | 16 | -2 | 24 | | | |

The results of partition function parameters calculations are in Tab. 1. The total accuracy of the computation was controlled by using of the total known number configurations, of the binomial coefficients (for 4x4 system there are 65536 configurations), also by other well-known parameters.

Other interesting outcome for given algorithm is in measuring of speedups which ones had super linear values for small number of processes. The running time of application was estimated with using of the library *time.h* and the MPI inline functions *clock(), MPI_Wtime(), ftime(),* system command "*time*" and duration writing of *datafile,* see Tab. 2.

TABLE 2. The measuring of process durations and speeding up by means of different methods for 5x5 lattice system. Time in sec.

| | | clock(); | MPI_ Wtime(); | ftime(); | System Command "time" | open-closing datafile |
|---|---|---|---|---|---|---|
| **The running time of the process #** | **0** | 37.910 | 37.9297 | 38.662 | 38.613 | 37.478 |
| | **1** | 37.460 | 37.4688 | 37.468 | | |
| | **2** | 37.710 | 37.7188 | 37.66 | | |
| | **3** | 37.530 | 37.918 | 38.612 | | |
| **Single process** | | 156.280 | 156.312 | 156.315 | 156.536 | 156.286 |
| **Speedup 4 nodes** | | 4.122 | 4.121 | 4.043 | 4.054 | 4.170 |
| **Speedup 8 nodes** | | 8.068 | 8.064 | 8.062 | 7.714 | 8.08 |
| **Speedup 16 nodes** | | 8.247 | 8.279 | 8.279 | 7.529 | 8.116 |

The cause of the strange and intriguing speedup values which ones are over maximal values (up to 4%) and in the contradiction with values predicted by the Amdahl's law, is not clear. The possible reason of such result could be quantities of the massive parallel logical-arithmetical operations over bit-vectors which ones could decrease because of cash hits are increases. The following decreases of the speedup are in connection with the growth of costs on message passing.

## 5. CONCLUSION

The high-performance parallel algorithm for rigorous calculation of partition function of the lattice systems of the finite number Ising spins was developed. The algorithm is scalable and it is possible to extend it for the running on huge calculation recourses. The facilities of calculation of 3D Ising lattices were elaborated. The nested iterations can be executed with utilization of the approach dynamical generation of the processes.

Unfortunately even with linear increasing of speedup by existing methods it is impossible to calculate $M_i$ and $E_i$ for each configuration of 10x10 system i.e. 100 spins in the given approach. So far as $2^{100}$ is approximately $10^{33}$, therefore for computational cloud with computer speed in order ExaMips ($10^{18}$ Mips) only for the summation until decillion it is necessary $10^{15}$ seconds or ~31,7 millions year! The same time needed for the 3D Ising lattice with structure 4x5x5. Of course, the availability of generating function could simplify the solution, but today we have not rigorous analytical solution of the simplest (!) ferromagnetism model even for 2D Ising lattice. The future approach to optimization of the chosen scheme of calculation of partition function could be done by means of taking into account the hypersymmetry of the this task. The solution obtained by authors can be used for comparison with experimental magnetization values for the finite atoms systems.

## 6. REFERENCES


[1] Istrail, Sorin (2000), "Statistical mechanics, three-dimensionality and NP-completeness. I. Universality of intractability for the partition function of the Ising model across non-planar surfaces (extended abstract)", **Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing**, ACM, pp. 87–96, MR2114521

[2] Barry A. Cipra, "The Ising model is NP-complete", **SIAM News**, 2000, Vol. 33, No. 6.

[3] N. Metropolis, S. Ulam The Monte Carlo Method, — **J. Amer. statistical assoc.** Vol. 44 № 247, 1949, pp. 335—341.

[4] Neumann **J., NBS Appl. Math, scries**, № 12, 1951, p. 36-38.

[5] Martin Niss, History of the Lenz-Ising Model 1920–1950: From Ferromagnetic to Cooperative Phenomena Archive for History of Exact Sciences, Vol. 59, №3, 267-318.

[6] Backster R., Exactly soved models in statistical mechanics", **M.: Mir**, 1985. – 488 p.

[7] Belokon V.I., Nefedev K.V., "Magnetic ordering in 1D and 2D models of finite number Ising spins systems", **Russian Physics Journal**, №3/2, 2010, pp. 19-23.

[8] Kittel Ch., "Statistical termodynamics". **M.**: 1997, 336 p.

[9] Brian W. Kernighan, Dennis M. Ritchie, "C Programming Language", **Published by Prentice-Hall in** 1988, ISBN 0-13-110362-81988.

[10] http://pelicanhpc.org/

# Implementation of FIX Engine and Order Management Systems using ASP.NET C#

Shweta Swarnkar
Department of Computer Science, Montclair State University
Montclair, NJ 07043, USA

and

John Jenq
Department of Computer Science, Montclair State University
Montclair, NJ 07043, USA

## ABSTRACT

Financial Information Exchange ("FIX") protocol, a public domain protocol, is a series of messaging specifications for the electronic communication of trade-related messages. FIX allows independently developed systems to communicate seamlessly and without licensing issues. It has been developed through the collaboration of banks, exchanges, broker-dealers, industry utilities and associations, institutional investors, and information technology providers from around the world. The vision of market participants is to standardize a common, global language for the automated trading of financial instruments. The objective of this project is to develop and implement a prototype FIX Engine and an Order Management System using Microsoft ASP.NET platform using C# programming language.

**Keywords**: FIX engine, Online Trading, ASP.NET

## 1.INTRODUCTION

Online Trading [1,2,3,4] is the buying and selling of stocks or other securities through the Internet. Instead of paying a broker to do the transaction for you, you type them in yourself. Brokerage fees are much lower, and transactions are completed more promptly.

Financial Information Exchange ("FIX") protocol [5, 6, 7], is a series of messaging specifications for the electronic communication of trade-related messages. FIX allows independently developed systems to communicate seamlessly and without licensing issues because FIX protocol is a public domain protocol. The FIX protocol has been developed through the collaboration of banks, exchanges, broker-dealers, industry utilities and associations, institutional investors, and information technology providers from around the world. These market participants share a vision of a common, global language for the automated trading of financial instruments.

A FIX engine is an application program which manages network connections, creates and parses communication messages both outgoing and incoming. A FIX engine also manages the session and application layers and is the single piece of software needed in order to FIX-enable trading or order management systems.

In the context of an online trading system, FIX engine serves as the interface to the outside world. Through connection of communication network, it connects you to the outside world and allows you to trade and exchange related information in a standard fashion. Thus, to FIX-enable an application refers to the integration of a FIX engine and connection to a routing network.

By maintaining electronic connectivity with trading counterparts, a FIX Engine sends and receives trading information and also monitors data integrity of the trading processes. The FIX engine actually is the component that handles low-level communications, and binds the protocol to your chosen programming language. A FIX Engine is a solution for companies who

plan to use the FIX protocol for electronic financial information exchange of financial data.

The Financial Interface exchange effort was initiated in 1992 by a group of institutions and brokers interested in streamlining their trading processes. FIX specifications are developed and managed by the members of organization known as "FIX Protocol Limited (FPL)". FPL is formed by major industry players.

The publication of the FIX standard in 1992 was followed by a number of products which can be categorized as first generation FIX engines. They were straightforward implementations of the FIX Protocol with their roots in traditional financial/accounting systems. They were typically built around a relational data base or on top of an application server. The most successful were Coppelia from Javelin [8], Financial Fusion from Sybase and Trinitech [9]. As FIX becomes more popular, the performance of these first generation solutions started to become an issue.

Second generation engines ran considerably faster than earlier versions, with Cameron FIX arguably being the first to market. Another significant second generation FIX engine which was the open source implementation, Quick FIX [10]. These FIX engines had raised the bar on performance but were still little more than straightforward implementations of the protocol.

The third generation FIX engine naturally implements the FIX protocol but it also serves as a platform for the processing of FIX messages. The FIX engine itself can be the most efficient and appropriate places to process those messages. Customers can configure in their own business logic for manipulating the FIX messages. FIX is now used by variety of firms and vendors. The latest FIX version is 5.0. Most of the current production versions are 4.1 to 4.4. FIX has become the de facto messaging standard for pre-trade and trade communication in the global equity markets, and is expanding into the post-trade space to support Straight through Processing, as well as continuing to expand into foreign exchange, fixed income and derivatives markets.

The FIX Engine can be combined with an Order management system that can store orders and operations in a database. In this project, the FIX Engine and an Order Management System for the Broker and Exchange is developed for online trading of stocks. This FIX engine is integrated with an Order Management System (OMS) for broker and exchange in order to generate and/or process FIX messages.

We outline of this report is as the following. Section two describes the system architecture and its implementation. Three-tier design was discussed. Section three list the system features. Section four concludes this report.

## 2. DESIGN AND SYSTEM IMPLEMENTATION

Three-tier design is adapted for our design of exchange server and OMS. These two servers exchange trading information through FIX engine. Any three-tier applications can be understood in terms of three different functional components: data management, application logic, and presentation. Different components are responsible for different purposes. The data management component is responsible for storing and managing all the persistent data used in the application. The presentation component is responsible for presenting the application data and functionality to the user and collecting the user's input to the application. The application logic component is the heaviest among all three. It implements the business logic of the whole application as well as addressing multiple functional and nonfunctional requirements by providing various services like caching, distributing computing, distributing transactions, data transformation, authentication, and many others.

As mentioned above, a FIX engine is integrated and link to an Order Management System (OMS) for broker and one FIX engine with exchange server. The Order management system which integrated with the FIX Engine will store orders and perform operations to the backend database. The Order Management System which links to the FIX Engine is used to generate and process FIX messages.

The FIX Engine establishes communication between broker's order management system and an Exchange's order management system. Orders placed by Broker OMS will be converted into FIX messages and transmitted by Broker's FIX engine to Exchange's FIX engine. The Exchange's OMS will receive all the orders after validation by an exchange's FIX engine. The FIX engine will also maintain TCP/IP and FIX sessions by generating heartbeat messages at regular time intervals. A Heartbeat monitors the status of communication link between broker's FIX engine and exchange's FIX engine.

This project is developed in Microsoft ASP.NET framework[11,12] and use of C# [13] language for development of the code. The information is stored in the Microsoft SQL Server [15] database management system. The web application is based on three tier architecture of software engineering. There is a clear separation of these three layers of the architecture. Most part of this web application is populated dynamically from the data stored in the database. This project uses a web server IIS server of Windows operating system.

## 3. SYSTEM FEATURES

### 3.1 Broker's Order management System

The home page of the Broker's Order Management system consists of a navigation menu that contains Home, Trade, My Account, Research & Ideas, Contact Us, About Us links. On clicking these menu links, it will open detail pages. User can access "Trade", "My Account", "Message Monitor" & "Message Archive" pages only after login to the application. Home page will contain "Login" and "Open An Account" submenu for user login. It will also have search buttons details by "Company Code" and "Company Name".

To open an account in the application user has to first create his user name and password. After submitting this information, a new page will open where user has to fill a form of his personal information like first name, last name, address, phone number, e-mail and his bank information. After submitting this form, user will get a message that account has been created successfully.

There is an Order Book for each user. User will be able to navigate to this page from Order Book link under the top navigation link Trade. This page will display order history of a user. This page will display company code of the company of which user has placed the order. It will also display quantity, order price, order date/time, side, open quantity, executed quantity and Time In force of the order. User can also modify or cancel his order before execution, if order status is active. **Figure 3.1** is a screen shot of an order book.

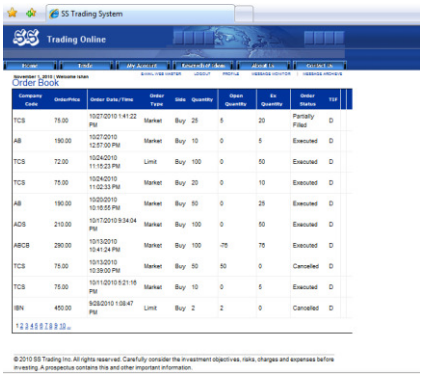



**Figure 3.1** An order book

User will be able to navigate to this page from Trade Book link under the top navigation link Trade. This page will display all the executed orders. This page will display company code, company name, quantity, tax, brokerage, total amount and total trade date/time.

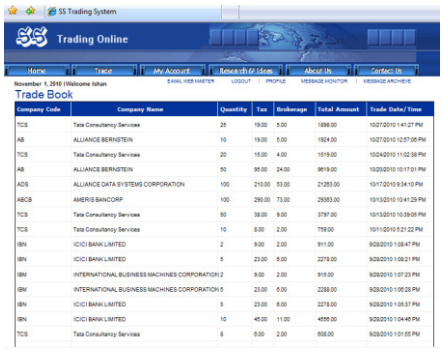



**Figure 3.2** A trade book

This Message Monitor page as shown in **Figure 3.3** will display after login in the broker's order management system. This page displays the entire message communication between Broker's and Exchange's FIX engines.

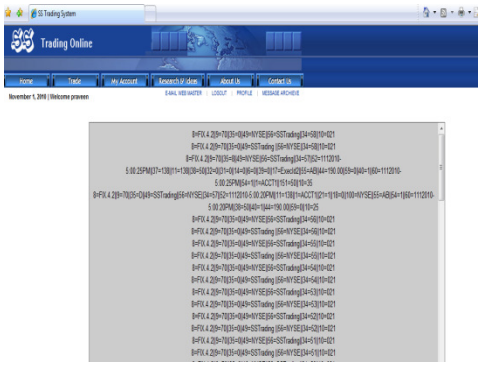



**Figure 3.3** FIX message monitor page

For illustration purpose, Figure 3.4 demonstrate a message which buying 5000 shares of IBM with $110.75 each share.

Each part is a set of fields. Body contains actual application message. Checksum is used to avoid transmission errors.tag 35 is very important tag that tells what type of information a message is carrying. Tag 8 indicates the version of fix protocol used to create the message. 35 = D new order single means the sender is sending a new order to trade the single security.

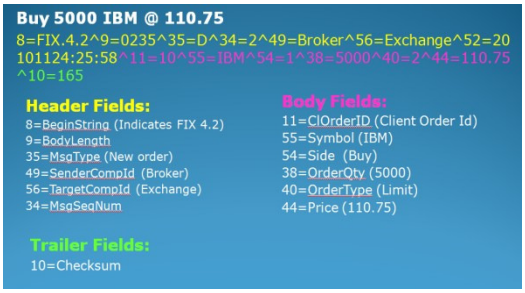



**Figure 3.4 Example of purchase of IBM stock**

There is a broker's Message Archive page displays messages generated on previous dates between FIX engines.

### 3.2 Exchange's Order Management System

We list some interesting pages for Exchange OMS.

The Order Monitor page will display orders placed from the Broker's side. This page will display Company Name, Order Id, Order Price, Order Date/Time, Order Type, Side Name, Ordered Qty, Last Share, Cumulative Qty, Leaves Qty, Order Status and TIF of the order. From this page Exchange can select Active or Inactive orders and will execute or cancel active order. This Page will also have submenus "Show Message" & "Logout" links.

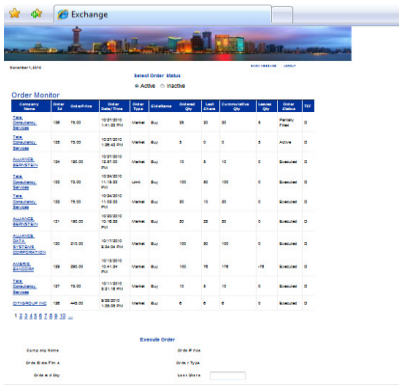



**Figure 3.5: Order Monitor Page Screen**

Similar to the broker System, the exchange system also can show message screen that displays after login in the exchange's order management system. This page displays the entire message communication between Broker's and Exchange's FIX engines. There are also Message archive for exchange server.

## 4. CONCLUSIONS and REMARKS

The benefits of a FIX engine include quick development and fast deployment for connectivity with new counterparties, easy and dynamic customization for maintaining and expanding the market reach. There are high flexibility and customizability in implementing the FIX Protocol and FIX variations. The efficiency of communication among financial

industry players is increased. As FIX engine utilizes FIX messages which reduces the time and complexity involved in connecting to multiple trading partners across different geographies. FIX engine will increase actual and potential completion between brokers for provision of trading services. Of course, it also reduces the explicit trading costs (e.g. brokerage commission and trading platform fees).

Over the years FIX has shown an extraordinary ability to evolve and adapt to the demands of its users, which has been key to its success. Its popularity has driven the evolution of FIX engine software. A FIX engine greatly simplifies connecting to exchanges or brokers. FIX engine implements all important aspects of the FIX protocol. FIX is a high performance, low-latency solution for order routing and market data.

Better FIX engines help to make the FIX Protocol more attractive to the market place. The result is a symbiotic relationship between the standard, and implementations of that standard, which bodes well for both. FIX is gaining increased attention within the Exchanges community as over three quarters of all exchanges surveyed supported a FIX interface. Currently, FIX is being used for equities, fixed income and foreign exchange trading for both cash and derivative products. FIX Engine is an enterprise-strength solution for companies who plan to use the FIX protocol for electronic financial information exchange of financial data, and Straight through Processing (STP).

Our implementation of FIX engine and OMS is very user friendly. More useful functionalities can be added in. As it will be a database application and will follow standard technologies, the maintenance of this application will be very easy. This application has lots of scope for future enhancements. One interesting area in FIX application is how to automate FIX message testing.

## 5. REFERENCES

[1] Online trading from All Business http://www.allbusiness.com/glossaries/online-trading/4946592-1.html

[2] Online trading from Answer.com http://www.answers.com/topic/online-trading

[3] How Online trading works http://money.howstuffworks.com/personal-finance/financial-planning/online-trading.htm

[4] Trading and broker related terms http://currencies.suite101.com/article.cfm/forex_online_trading

[5] Financial Information Exchange Wikipedia http://en.wikipedia.org/wiki/Financial_Information_eXchange

[6] What is Fix http://fixprotocol.org/what-is-fix.shtml

[7] FIX protocol Specifications http://www.fixprotocol.org/specifications/

[8] Javelin Technologies http://www.javelin-tech.com/main/index.html

[9] Financial Fusion of Sybase http://www.sybase.com/solutions/financialservicessolutions/banking

[10] Quick Fix: http://www.quickfixengine.org/

[11] Official ASP.NET web site http://www.asp.net/

[12] Zak Ruvalcaba, Build Your Own ASP.NET Website Using C# & VB.NET, Site Point Pty.Ltd. 2009

[13] Paul Deitel *Visual C# 2008 How to Program*, Pearson Publishing

[14] Christinan Nagel, Bill Evjen, Jay Glynn, Morgan Skinner, Karli Watson, Allen Jones, *Professional C# 2005*, Wiley Publishing Inc.

[15] Microsoft, *SQL Server 2005 Books Online*, http://msdn.microsoft.com/en-us/library/ms130214(SQL.90).aspx

# Quantum Energy Teleportation: New Energy Distribution Inside Quantum Couputer Without Thermal Decoherence

Masahiro Hotta[1] *

[1] *Graduate School of Science, Tohoku University*
*6-3 Aza Aoba, Aramaki , Aoba-ku, Sendai, Japan.*

## 1 Abstract

Recently it is proven that energy can be teleported by a new quantum protocol

By use of ground-state entanglement, rapid energy transportation can be achieved by local operations and classical communication before heat generation. This offers a new scheme of energy distribution in quantum computers without thermal decoherence.

## 2 Description of QET

Recently it is proven that energy can be teleported by a new quantum protocol called quantum energy teleportation (QET for short) [1]-[7]. This QET protocol retains all physical laws including local energy conservation. At first, we measure zero-point fluctuation of a subsystem A of a many-body system in the ground state. The post-measurement state is not the ground state in general. Therefore positive amount of energy $E_A$ must be injected into A by the measurement device. This energy is regarded as energy input of the teleportation. After the measurement, the result is announced to another subsystem B at a speed much faster than energy-carrier velocity. By performing an operation dependent on the result to B, we can extract energy $E_B$ from B, which is regarded as energy output of the teleportation. By extracting positive energy $+E_B$ to outside systems, the zero-point fluctuation at B of the channel can be more suppressed than that of the ground state, yielding negative energy $-E_B$ at B. Here we fix the origin of the energy density of the channel such that the expectational value vanishes for the ground state. Thus the total energy of the channel is nonnegative. In general, quantum interference among total energy eigenstates can produce various states containing regions of such negative energy density, although the total energy remains nonnegative. QET increases not the total energy at the region of B but instead the percentage of available energy at the region of B to be harnessed for arbitrary purposes by decreasing the zero-point energy of B.

This protocol is expected to offer a new scheme for energy distribution inside quantum computers without heat generation. This enables us to execute quantum tasks using the distributed energy before heat destroys quantum coherence.

Physical energy carriers do not play any role for the energy extraction during this short-time QET process. Soon after a one-round completion of the protocol, the input energy $E_A$ still exists at A because late-time evolution of the energy carries does not begin yet. Let us imagine that we attempt to completely withdraw $E_A$ by local operations at A after the extraction of energy from B. If this was possible, the energy gain $E_B$ might have no cost. However, if so, the total energy of the channel became equal to $-E_B$ and negative. Meanwhile, we know that the total energy of the system is nonnegative by our definition of the origin of the energy density. Hence, it is not allowed physically to withdraw energy larger than $E_A - E_B$ by local operations at A. This argument also implies that $E_B$ is upper bounded by $E_A$. Another reason for this inability of complete extraction of $E_A$ is because the first measurement made at A breaks the ground-state entanglement between quantum fluctuation at A and quantum fluctuation at B. Therefore, after the measurement at A, the ground state (zero-energy state) is no longer recovered only by A's local operations, which do not restore the above broken entanglement. Hence it can be concluded that a part of input energy $E_A$ cannot be extracted from A during the short time scale. QET enables this residual energy at A to be effectively extracted in part as $E_B$ from the distant point B by use of the measurement information of A. Needless to say, after the completion of the QET process, the positive energy $E_A$ at A compensates for the negative energy $-E_B$ at B during the late-time evolution of the energy-carrier dynamics. Though heat can be generated in this late-time period, heat does not arise in intermediate subsystems between A and B in the early-time period of QET.

The QET protocols can be implemented, at least theoretically, to various physical systems, including spin chains, cold trapped ions, harmonic chains, quantum fields and quatum Hall edge current. The reference list of QET can be seen in [8].

## References

[1] M. Hotta, "A Protocol for Quantum Energy Distribution", **Physics Letter,** A372, 2008, pp5671.

*hotta@tuhep.phys.tohoku.ac.jp

[2] M. Hotta, "Quantum Measurement Information as a Key to Energy Extraction from Local Vacuums", **Physical Review,** D78, 2008, pp045006.

[3] M. Hotta, "Quantum Energy Teleportation in Spin Chain Systems", **Journal of the Physical Society of Japan,** Vol. 78, 2009, pp034001.

[4] M. Hotta, "Quantum Energy Teleportation with Trapped Ions", **Physical Review**, A80, 2009, pp042323.

[5] M. Hotta, "Controlled Hawking Process by Quantum Energy Teleportation", **Physical Review**, D81, 2010, pp044025.

[6] M. Hotta, "Quantum Energy Teleportation with an Electromagnetic Field: Discrete vs. Continuous Variables", **Journal of Physics A: Mathematical and Theoretical**, Vol. 43, 2010, pp105305.

[7] M. Hotta, "Energy Entanglement Relation for Quantum Energy Teleportation", **Physics Letter**, A 374, 2010, pp3416.

[8] M. Hotta, "Quantum Energy Teleportation: An Introductory Review", 2011, extended version, http://www.tuhep.phys.tohoku.ac.jp/~hotta/extended-version-qet-review.pdf.

# Approach of using Ultra-Wideband-Radio in Industrial Real-Time Ethernet Networks

**Kristian Stieglitz, Joerg F. Wollert**
**Department of Electrical Engineering and Computer Science,**
**University of Applied Sciences**
**Bochum, 44801, Germany**
**{kristian.stieglitz, joerg.wollert}@hs-bochum.de**

## ABSTRACT

Within the last years, the utilization of wireless solutions for industrial applications has become very popular. Especially in harsh environments, mobile and rotating scenarios, or at positions difficult to access, the advantages of radio technologies are obvious. In addition to that, there is a great potential on saving time and money during planning, installation, and commissioning of plant sections. However, due to the lower capacity and reliability of wireless links compared to wired ones, time critical domains like factory automation or motion control, can hardly be served by radio based solutions. Meanwhile, Ultra Wideband (UWB) a wireless transmission technology for high bitrate more than 480 Mb/s and short range (1-10m), offer opportunities, to meet the requirements of high speed industrial communication systems.

The full paper gives a short overview of important industrial Real-Time-Ethernet (RTE) technologies and the ECMA-368 UWB radio. Furthermore, we present the relevant performance characteristics of ECMA-368 and compare them with the requirements of current RTE-Systems. The paper concludes with the approaches to realize typical use cases of wireless solutions in industrial applications.

**Keywords**: Real-Time-Ethernet, wireless, ultra wideband, ECMA-368, automation, motion control, RTE-Bright, WSAN

## 1. INTRODUCTION

Short Range WPAN Networks like Bluetooth or Zigbee gained significant popularity in recent years. Few domains of industrial automation already benefit from using wireless applications. Especially in mobile and rotating scenarios, harsh environments, or at positions difficult to access, the advantages of radio technologies are obvious. In this use cases radio links often offer a lower error probability compared to sliding contacts and cable connections. Specific mobile scenarios are only realizable, based on radio technologies. Furthermore there is a high potential on reducing time and money during planning, installation, and commissioning of plant sections. It has been estimated that common wiring costs in industrial installations are about US $ 130,..,650 per meter and adopting wireless solutions could save 20 %,..,80 % of these costs [1].

Besides the development of radio-based systems for the industrial use, in the last years classic fieldbus systems for industrial automation are slowly been replaced by Real-Time Ethernet fieldbus systems. Current RTE offer significantly shorter cycle times with much higher data rates. Because of this features RTE are very good suitable for the use in factory automation and motion control. However, this performance exceed by far the capabilities of current industrial wireless solutions, which are based on low band wide radio system such as IEEE 802.11/WLAN [2], IEEE 802.15.1 [3] / bluetooth [4] or IEEE 802.15.4 [5].

Further, due to the fluctuating nature of these radio communication channel, the mentioned time critical low latency domains, can hardly be served by radio based solutions [6]. Particularly in industrial environments, with several moving and metallic obstacles, reliable wireless data communications are very difficult to obtain.

In this context, the Ultra Wideband (UWB) radio technologies offer opportunities, to meet the requirements of high speed industrial communication systems. First standards for Ultra-Wideband (UWB) radio technologies have been published in 2006 and 2007. In general, UWB offers the following advantages:

- Low latency times, due to extreme short symbol durations, what additionally offers the possibilities for precise ranging.
- Robust against the effects caused by multipath scattering. Reflection and scattering are frequency selective. Using a high bandwidth reduces the probability of deep fadings over the whole frequency range.
- Energy efficiency, due to the low spectral density power.

Especially the first two statements underline the potential suitability of UWB for low latency real time applications. Possible use cases are:

- Cable replacement in high speed real time Ethernet (RTE) connections (RTE-Bridge).
- Application in wireless sensor/actor networks (WSAN).

ECMA-368 is one of the current available UWB technologies. It defines high data rate UWB and is based on the specifications of the WiMedia Alliance. ECMA-368 is the foundation for a bundle of commercial protocols, like WiNET for TCP/IP support and Certified Wireless USB. Analyses of the ECMA-368 specification as well as practical tests have shown that the standard is suitable for the realization of the mentioned use cases [7].

This paper is published within the scope of the project „Ultra Wideband Interface for Factory Automation" (UWIfac), funded by the German ministry for education and research (BMBF).

The rest of the paper is organized as follow, Chapter two and three gives a short overview of the current industrial RTE-technologies and the ECMA-368 UWB radio and present the relevant performance characteristics of them. In the following Chapter possible applications for UWB Radio in RTE networks would be presented and discussed. The paper concludes with presentation of the used hardware and software framework.

## 2. Modern Industrial Communication Systems

The traditional fieldbuses and the directly coupled signals are slowly being replaced by Ethernet based communication solutions [8]. The move towards Ethernet as the basic communication platform is mainly based on the efficient price/performance relationship of the technology. Corresponding to Figure 1, the performance of real time Ethernet (RTE) protocols has historically evolved into three generations [9].

The protocols using the whole Ethernet TCP/IP stack and on top a real time specific application layer belong to the first generation. Good examples are Ethernet/IP [10] and Mod-Bus/IDA [11]. Since the whole TCP/IP protocol stack is used, the real time performances are limited. Guaranteed update times of about 100 ms can typically be reached.

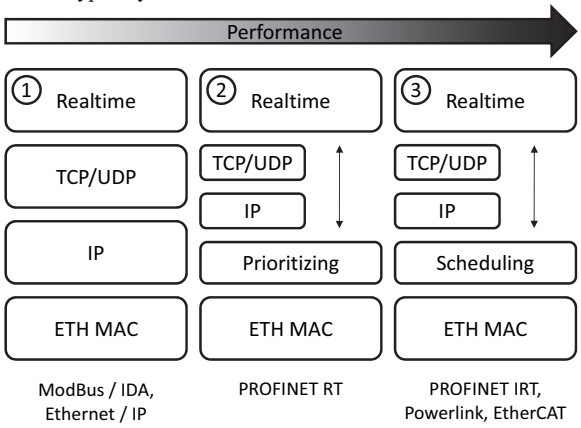



**Figure 1: Classification of industrial-Ethernet Protocols[8]**

Protocols of the second generation are a tradeoff between native Ethernet standard versus achievable real time performance. The transport and network layers are bypassed in order to achieve a more efficient real time communication. By means of this optimization, update times in the range of about 10 ms can be reached. A good example is PROFINET RT [12].

Protocols which are changing/replacing the original MAC scheme are part of the third generation. For those protocols, specific hardware or software is necessary. Good examples are Ethernet Powerlink [13], EtherCAT [14] and PROFINET IRT[12]. Depending on the protocol, summation or individual frames and cut-through operations are used, in order to exchange process date at extreme low latency times. The update times of third generation RTE protocols reach down to below 250 µs, according to the number of nodes joining a network.

The interested protocols of the third generation roughly can be divided into three functional principles:

**Token-based methods:** These include the RTE EtherCAT and SERCOS III. In this principle the master note transfers on or more Ethernet frames in a row. These Ethernet frames carry all process data of the slaves. This process data are passing each slave and the last slave sent them back to the master. Physically, the slaves are connected in a ring or linear topology. The process data in the Ethernet frames are processed by the slaves "on the fly". So the Ethernet frames are constantly delayed less than 500 ns. Important for this principle is the need of a full duplex communication.

**Request Response Procedures:** This includes the RTE PowerLink. In this principle the process data between the master and the slave where transferred in a separate Ethernet frame for each slave. The Master requests the slaves to take over the process output data and to send their input process data back to the master. This method does not require am full-duplex communication.

**Consumer Producer process:** This includes the RTE PROFINET. In this principle the master determined for each slave at the beginning, when the process data have to be transmitted. The transmission timings are optimized to the network, so that the collisions between the slaves are minimized. In real-time operation, the slaves transmit there process data periodically without further requests by the master. Also the master sends its output process data network optimized to the slave notes. This process requires also a full duplex communication.

| | | Powerlink | PROFINET IO/IRT | SERCOS III | EtherCAT |
|---|---|---|---|---|---|
| **Real time Layer** | | | | | |
| Real Time over Transport Layer | | - | - | - | - |
| Real Time over ETH-MAC | | x | - | - | - |
| Modified MAC | | - | x | x | x |
| Specific Phy Layer | | - | - | - | - |
| **Transmission Times** | | | | | |
| IAONA-Echtzeitklasse | | 4 | 4 | 4 | 4 |
| minimale Zykluszeit [s] | | 200 µs | 250 µs | 31 µs | 30 µs |
| Obergrenze des Jitters [s] | | < 1 µs | < 1 µs | < 1 µs | < 1 µs |
| **Strategies und Synchronization** | | | | | |
| Introduction of TDMA | | x | x | x | - |
| Token based Method | | - | - | x | x |
| Introduction of prioritization | | - | x | x | - |
| Response Request Method | | x | - | - | - |
| On the Fly Frame Processing | | - | - | - | x |
| Synchronisation über Sync-Signal | | x | - | x | - |
| Time Synchronization (like IEEE 1588/PTCP) | | x | x | - | x |
| **Topoloegie & Verteiller** | | | | | |
| Logic Topologies: | Tree | - | x | - | - |
| | Line/Bus | x | - | - | - |
| | Ring | - | - | x | x |
| Physical Cabling: | Baum | x | x | - | x |
| | Line/Bus | x | x | x | x |
| | Ring | - | x | x | x |
| Distributor: | Hubs | x | - | x | - |
| | Switches | - | x | - | x |
| | Special | - | x | x | x |
| **ETH-Standards Conformance** | | | | | |
| Standard ETH-Frames Transmission | | x | x | x | x |
| ETH-Standard Allowed in Higher Layer | | x | x | x | x |
| Capping is Compiant with the ETH-Standard | | x | x | x | x |
| Switches are Compliant with the ETH-Standard | | x | - | x | - |
| TCP/UDP/IP-Frames in Network Allowed | | x | x | x | - |
| nRT-Traffic in Specific Time Slots | | x | x | x | - |
| Standard PCs in Network Allowed | | - | x | - | - |
| Cross Traffic Allowed | | - | x | x | x |
| Ether Type | | 0x88AB | 0X8892 | 0x88CD | 0x88A4 |

**Figure 2: Properties of industrial-Ethernet Protocols**

The table in figure 2 shows the performance and transmission

characteristics of the mentioned RTE. The figure 3 shows the theoretical possible minimum cycle times for the RTEs depending to the number of slaves for different sizes of process data. Practical the following minimum cycle times are currently available:

- 250 ms Profinet IRT
- 200 ms Powerlink
- 50 ms EtherCat
- 32 ms SERCOS III

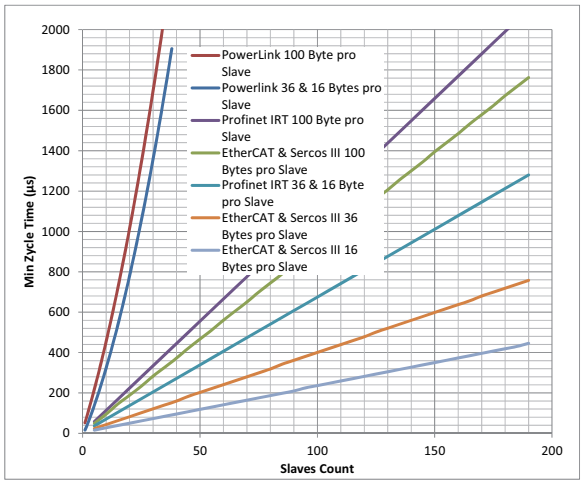



**Figure 3: Updatetime of industrial-Ethernet Protocols**

## 3. ULTRA-WIDEBAND

The first regulation for UWB devices within a frequency range between 3.1 GHz and 10.6 GHz was published by the FCC in 2002 [15]. The maximum e.i.r.p. power density is limited to 41.3 dBm/MHz. More restrictive regulations concerning the frequency ranges, channel occupation, and maximum power spectral densities followed for Europe, Japan, Korea, and China since 2007.

UWB follows the approach of a parallel utilization of the frequency spectrum with a large band-width and a low spectral density power, hence being immune and appearing as noise to coexisting narrow band technologies [16].

Based on the specifications of the WiMedia Alliance [17], the first UWB standard ECMA-368 [18] was published in late 2006 and is available in version 3.0 since 2008. It uses a Multiband OFDM (MB-OFDM) scheme and supports high datarates of up to 480 Mb/s on the PHY layer. The amendment IEEE 802.15.4a [19] is the second standard for UWB PHY and MAC layers and was published in 2007. The standard uses Direct Sequence UWB, bursts of impulses, to generate signals and aims at a ultra low power and low datarate communication (100 kb/s to 27 Mb/s) with precise ranging capabilities. However, the second standard is unattractive for the use in RTE networks, due to its low data rates.

### *ECMA-368*

ECMA-368 defines high data rate UWB PHY and MAC layers and is based on the specifications of the WiMedia Alliance. The standard builds the foundation for a set of commercial protocols, like Certified Wireless USB (CW-USB) [20], and WiNET for TCP/IP support. The standard defines fourteen 528 MHz frequency bands, which are split up into six band groups. Each band group consists of three bands with the exception of band group five, which consists of only two bands.

Data is encoded using MB-OFDM with 122 sub-carriers (100 data, 10 guard, 12 pilot). The different data rates are realized, using a convolutional code with coding rates of 1/3 to 3/4, time domain spreading (TDS), and frequency domain spreading (FDS). TDS makes use of time diversity, by redundantly transmitting data over two consecutive symbols. FDS makes use of frequency diversity, by redundantly transmitting data over two OFDM carriers. Depending on the data rate, distances between three to ten meters are achieved at an output power of -41.3 dBm/MHz. This is a sufficient range for applications in typical production cells.

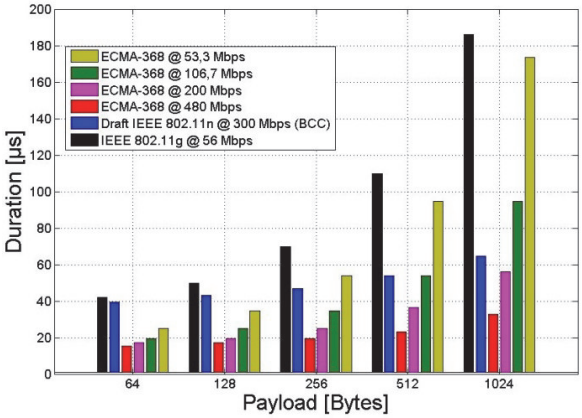



**Figure 4. Comparison of the physical transmit durations between ECMA-368, IEEE 802.11g and IEEE 802.11 Draft n [7]**

Figure 4 shows a theoretical comparison of the physical transmission duration of ECMA-368 for different datarates, IEEE 802.11g at 56 Mb/s, and IEEE 802.11 Draft n at 300 Mb/s (greenfield) for different amounts of payloads. Because of the shorter protocol overhead, even at 53.3 Mb/s ECMA-368 outperforms IEEE 802.11 Draft n for payload length of > 100 bytes. Thereby, the energy consumption of ECMA-368 is approximately one magnitude lower compared to IEEE 802.11.

The standard specifies a fully distributed MAC layer with no explicit coordinating device. A transmission channel is divided into superframes, each consisting of 256 medium access slots (MAS). Each MAS has a duration of 256 µs. A superframe is composed of a beacon period (BP) at the beginning and a following data period. The BP constists of a variable number of beacon slots with a duration of 85 µs. Within the beacon slots, devices exclusively send beacon frames, containing information on the device. With each device joining a network, called beacon group, the beacon period gets extended by one beacon slot. Two beacon slots are always reserved for devices to join in the future. Beacons are also used to negotiate and coordinate the channel access among the device within the data periods of superframes. The standard defines the two access methods distributed reservation protocol (DRP) and prioritized contention access (PCA).

The DRP gives exclusive access to reserved blocks of MASs for a reservation target and a reservation owner, where the target is a single device and the owner may also belong to a multicast group. Data transmission is always initiated by the target device. Within blocks of non-reserved MAS, devices may access the medium using PCA. PCA is a CSMA/CA scheme.

In summary it can be said that the investigations presented by RTE and ECMA 368 have shown that the realization of the presented use cases should be possible with cycle times well below 1 ms. However, in order to get full access to the PHY capabilities,

the following weak points of the MAC have to be overcome:

- Depending on the number of devices, participating in a beacon group, the length of the BP varies between 512 µs and 8.16 ms. At these times, no data transfer is allowed.
- In principle, DRP offers a time synchronised access to the medium, suitable for deterministic real time protocols. But the duration of 256 µs of a MAS is alligned with respect to large payloads. When using frames with small payloads, like it is common for industrial protocols, about 80 % of this time would be left unused.
- PCA gives more flexibility for devices to access the channel. But contention aware protocols do not guarantee a deterministic channel access.

## 4. Industrial Use Cases

Within the scope of the project „Ultra Wideband Interface for Factory Automation" (UWIfac), funded by the German ministry for education and research (BMBF), the feasibility of the following two use cases should be proved:

- **Time synchronized wireless sensor/actuator Network (WSAN):** This is understood as a real-time Ethernet slave who is able to exchange the completely process data of its attached wireless sensors and actuators within a real-time Ethernet cycle (typically 100 µs to 1 ms).
- **Real-time Wireless Bridge (RTE-Bridge):** Descripts a wireless point-to-point connection with an extremely reliable almost delay and jitter free communication behavior within the permitted real time limits.

***Time-synchronized wireless sensor/actuator Network:***

The requirements to the UWB radio technology are different for both use case. In the use case of WSAN, the RTE is not affected by the wireless technology, because the gateway note is a RTE-Slave of the respective RTE itself. This gateway has its own process data, which stores the input and output process data of the individual radio sensor/actuator notes. Regardless of the RTE-cycle these process data are exchanged between the gateway and the wireless nodes in its own sub cycle. This is a master-slave relationship with a request response communication behavior. For the RTE network the WSAN gateway note looks like an RTE-coupler with a sub-bus. Hence, the requirements for this use case are:

- a minimum cycle time of the sub-bus
- a very reliable radio link

The minimum cycle time of the sub-bus depends on the number of wireless notes. To keep the cycle time small for many wireless devices, the wireless devices can be subdivided to multiple smaller groups each with their own gateway note. Therefore, ECMA-368 offers the possibility to operate multiple systems simultaneously in different radio channels. This parallel operation is very limited for conventional radio systems. In principle, this use case can be realized functional in any case. However, it needs to be tested, if the desired performance can be achieved.

***Real-time Ethernet Wireless Bridge:***

For the use case RTE-Bridge the requirements to the UWB radio technology are much tougher than to the previous use case. The system affects directly the RTE communication, because it practical replaced a cable connection between two slave-RTE or a network switch or hub. Especially for the RTE of the third generation the demands for a constant time behavior of these network components are very high, to achieve an exact synchronization of the RTE slaves. Assuming that the RTE-Bridge is replaced an Ethernet cable, three requirements are important for the RTE-bridge:

- Full duplex communication
- Jitter- and nearly delay-free transmission of Ethernet frames
- High reliability against loss and damage of the Ethernet frames

It is obvious, that wireless technologies cannot reach these characteristics of a cable. Therefore it is important to find out, how high the requirements for the specific RTE really are.

The full duplex communication is one of the most critical requirements for conventional radio systems. Conventional radio systems are based on a single channel half-duplex communication. That means, when RTE- system requires a full duplex communication, a single-channel radio system wouldn't meet this requirement. RTE which needs a full duplex communication are, SERCOS III, EtherCAT and PROFINET RT / IRT.

Thus, a radio system with full duplex communication has to be designed with two channels, so that each transmission direction has its own radio channel.

The next requirement is the delay and jitter free forwarding of Ethernet frames. Conventional Ethernet wireless bridges operate on the "store and forward" principle. According to this principle, also commercially available Ethernet switches are working. In this case, the incoming Ethernet frame is received completely and temporarily stored before being forwarded. The Advantage of this method is that damaged or unwanted Ethernet frames can be discarded. The Disadvantage of this method is that the Ethernet frame is delayed by at least the time of receiving the whole frame. Thus, the forwarding delay is variable. For radio-based RTE-Bridges this effect is doubled, since the "store and forward" principle is used on the Ethernet and Radio interface. Therefore almost all RTE systems of the third generation only allow hubs or switches how operates according to the "cut through" principle. This principle forwards the Ethernet frame after the first bytes are received. This principle has the advantage that the delay time is not dependent on the Ethernet frame length and thus the forwarding jitter is very low. Therefore the RTE-Bridge should operate according to the "cut through" principle. This means that the RTE-Bridge starts to forward the data to the radio interface after receiving the first bytes of an Ethernet frames. On the receiving side an Ethernet frame is transmit as soon the first bytes arrive on the radio interface.

The next requirement is the high reliability against lose or data corruption of the Ethernet frames. Ethernet frames can be lost or damaged, that applies to radio as well as to cable connections. However, radio connections are not so reliable compared to cable-connections. Since the RTE systems expect the reliability of a cable connection, the RTE Bridge must also satisfy these demands.

RTE systems do not retransmit damaged or lost Ethernet frames again, because the data would be already obsolete then. Instead, RTE systems tolerate a certain number of lost or corrupted Ethernet frames. Thus, the RTE-Bridge has to be at least so reliable that the tolerable number of lost and corrupted Ethernet frame is not exceeded. ECMA-368 already offers a very reliable radio communications.

## 5. Approach Discussion

In the following chapter, three possible solutions for the use case RTE-Bridge would be presented and discussed:

- Transparent RTE-Bridge
- RTE specifically optimized RTE-Bridge
- RTE-Bridge as a gateway between two independent RTE networks

**1) Transparent RTE-Bridge:**

This approach would provide the best solution. It has the advantage that the entire Ethernet frame is transmitted without any modification so there is no need to implement higher RTE specific protocol layers. However, in this approach the RTE-Bridge have to come the characteristics of an Ethernet cable especially close, because no RTE specific optimization can be done, which would relax the high requirements. These Requirements are:

- Two-channel full-duplex communication over the UWB radio
- "cut through" behavior for the radio and Ethernet interface with constant transmission delays.

The biggest challenge would be to achieve the constant and small forwarding delays, which are needed to get an accurate synchronization of the RTE network.

**2) RTE specifically optimized RTE-Bridge:**

In this approach the RTE-Bridge would be implemented for each RTE with specific optimization. Depending on the type of RTE it can be decided, whether the high demands of an Ethernet cable can be reduced. In the case of the RTE Powerlink, for example, no full-duplex communication is needed, so that the RTE-Bridge can be realized with a single channel UWB radio. Since in this approach the RTE-Bridge implements the RTE protocol stack, only the necessary process data are needed to be transmitted over the radio interface. That means that the protocol overhead of the Ethernet link layer and the RTE Layer could be cut off for the transmission over the radio. Also the bridge endpoints can be synchronized to the RTE networks cycle time so that the RTE network timings could be corrected by the RTE-Bridge to keep the jitter small. The disadvantage of this approach are the high implementation efforts, as each RTE needs an own implementation.

**3) RTE-Bridge as a gateway between two independent RTE networks:**

In this approach the process data are exchanged between two independent RTE networks. From the perspective of RTE both radio nodes of the RTE-Bridge are implemented as a RTE slave of the connected RTE network. Both nodes of the RTE-Bridge are exchanging independent from their RTE network cycle time, their own set of input and output process data, so the RTE-Bridge represent a gateway between both RTE network. It is the task of the master controllers in both RTE networks to selected suitable process data for the exchange between the RTE-Bridge notes. The implementation of the RTE-Bridge nodes is RTE dependent. The advantage of this solution is that the RTE networks and their timing aren't affected trough the RTE-Bridge similar to the WSAN use case. Thus, there are no synchronization problems in the RTE network. Since, with this approach the process data can be transmitted independent from the RTE protocol, two different RTE systems can be connected with each other. Therefore, the endpoints of the RTE-Bridge have to be implemented specific to the selected RTE. The Disadvantages of this solution is that two RTE networks must be setup with their own master. Also the endpoints of the RTE-Bridge have to be implemented for every supported RTE protocol. As the exchanged process data have to be processed by both master nodes before they could be transferred to the I / O, the performance potential of this approach is smaller than of the previous approaches. Furthermore, the master nodes are responsible for the timing and synchronization between both RTE networks. Similar to the WSAN use case, in principle this approach can be realized functional in any case. It also needs to be tested, if the desired performance could be reached.

## 6. HARDWARE AND SOFTWARE

For the prototype development evaluation boards called Anaxo are used. This evaluation board, showed in figure 5, was developed from the Institute of Embedded Systems (InES) of the Zurich University of Applied Sciences (ZHAW) [21, 22]. The board contains an Altera Cyclone III FPGA with four 10Base-T/100Base-TX Ethernet interfaces. The WiMedia compliant UWB module on the evaluation board contain the ECMA-368 UWB PHY RTU7012. This PHY is supporting the band groups 1 and 3 with all data rates from 53.3 Mbps to 480 million Mbit/s [21]. Through the use of an FPGA, other components like a soft-core, the UWB MAC IP and a slave-RTE IP could be programmed in the FPGA without changing the hardware. The used ECMA 386 UWB MAC is a VHDL-IP. This UWB MAC was also developed by InES. Since the UWB MAC has been developed in VHDL, it can be easily optimized to the desired RTE applications.

The UWB MAC can independently read/write data directly from the memory and transmit/receive them over the radio interface. No microprocessor is needed for data processing, which leads to very short transmit and receive delays. For the use with RTE, an Ethernet interface could be implemented in this way that the incoming and outcoming Ethernet data were also read/write directly to the transmission/reception memory of the UWB MAC. Such an implementation could be a possible solution for the first approach of the RTE-Bridge use case. For the use cases WSAN and RTE-Bridge, in the RTE specific variants two and three, commercial RTE-Slave-IPs are available, which can be used for the RTE slave implementation on the FPGA. For the implementation of these approaches a RTE-Slave-IP could be programmed together with the UWB MAC IP into the FPGA and the UWB MAC could be configured in this way, that the MAC could access directly the input and output process data memory of the RTE-Slaves-IP. Since this solution also don't need microprocessor for data processing, an optimum response and delay time could be achieved.

Since EtherCAT is one of the most powerful RTE [23] and a commercial IP for the selected FPGA is available, the presented concept will be tested with this RTE first. If the desired goals could be achieved with this RTE, it can be assumed that this concept is portable to other RTE protocols. The porting effort to other RTE protocols would be low, due to the chosen hardware platform.

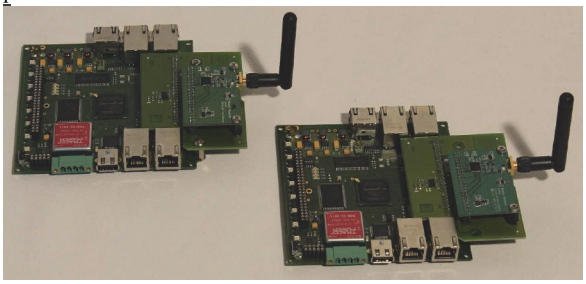


**Figure.5. Anaxo – Ultra Wideband Communication Development Kit [21]**

## 7. CONCLUSION

As a whole, one can say ECMA 368 PHY has the capabilities, to achieve a reliable wireless data transfer beyond the timelines given by state of the art industrial wireless solutions. However, without modifying the MAC, it will be hardly possible to get full access to these capabilities. Thanks to the presented hardware platform, these modifications are possible, so the UWB MAC could serve the needs of industrial high-speed RTE communication. For the presented use cases WSAN and RTE-Bridge the requirements for the UWB radio technology were investigated and several possible solutions discussed. In principle, with the chosen hardware platform a functional implementation of the use cases should be possible. With the limited range of about 10 m ECMA-368 is well suited for the local operation within dedicated production cells. With concern to coexisting technologies, the short range actually is of advantage for this use case.

Within the scope of the project „Ultra Wideband Interface for Factory Automation" (UWIfac), funded by the German ministry for education and research (BMBF), further theoretical and practical evaluations of the presented approaches, will be accomplished. Measurements regarding latency, error, and coexistence properties will be accomplished and the results will be published in the future.

## REFERENCES


[1] Chehri, A.; Fortier, P.; Tardif, P.: An Investigation of UWB-Based Wireless Networks in Industrial Automation, proceedings of IJCSNS col. 8, no. 2, February 2008.

[2] LAN/MAN Standards Committee of the IEEE Computer Society. Information technology – Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Networks – Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Revision of 2007.

[3] LAN/MAN Standards Committee of the IEEE Computer Society. IEEE Standard for Information technology – Telecommunications and information exchange between systems – Local and metropolitan area net-works – Specific requirements – Part 15.1: Wireless Me-dium Access Control (MAC) and Physical Layer (PHY) Specifications for Wireless Per-sonal Area Networks (WPANs).

[4] Bluetooth, http://www.bluetooth.com, 2009.

[5] LAN/MAN Standards Committee of the IEEE Computer Society. IEEE Standard for Information technology – Telecommunications and information exchange between systems – Local and metropolitan area net-works – Specific requirements – Part 15.4: Wireless Me-dium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wire-less Personal Area Networks (LR-WPANs), Revision of 2006.

[6] Vedral, A.; Wollert, J. F.; Buda, A.; Altrock, R.: The Capability of Bluetooth for Real-Time Transmission in Automation, proceedings in IASTED Network and Communication Systems (NCS 2006), March 2006

[7] Buda, A.; Stieglitz, K.: Using Wireless Systems in Realtime Ethernet Fieldbuses – Problems and Solutions XI International PhD Workshop OWD 2009

[8] Ethernet-Based Device Networks Worldwide Outlook. Market analysis and forecast through 2012, ARC Advisory Group 2008.

[9] Jasperneite, J.; Schumacher, M.; Weber, K.: Limits of increasing the performance of industrial Ethernet protocols, proceedings of ETFA 2007, pp. 17-24, 2007.

[10] Ethernet/IP, Ethernet Industrial Protocol (Ethernet/IP), www.Ethernet-ip.org, 2009.

[11] ModBus, www.modbus.org, 2009.

[12] PROFINET, PROFIBUS International, http://www.profinet.com, 2009.

[13] Powerlink, Ethernet Powerlink Standardization Group, www.Ethernet-powerlink.org, 2009.

[14] EtherCAT, Ethernet Technology Group (ETG), www.ethercat.org, 2009.

[15] Federal Communications Commission FCC: 02 48A1 Revision of Part 15 of the Commission's Rules Regarding Ultra-Wideband Transmission Systems, February 2002, revision of 2007.

[16] Nekoogar, F.: Ultra-Wideband Communications-Fundamentals and Applications, Prentice Hall Communications Engineering and Emerging Technologies Series, ISBN: 0-13-146326-8, 2005.

[17] WiMedia Alliance, http://wimedia.org, 2009.

[18] Ecma International: Standard ECMA-368: High Rate Ultra Wideband PHY and MAC Standard, 3rd Edition, 2008

[19] LAN/MAN Standards Committee of the IEEE Computer Society. IEEE Standard for Information technology – Telecommunications and information exchange between systems – Local and metropolitan area networks – Specific requirements – Part 15.4a: Wireless Me-dium Access Control (MAC) and Physical Layer (PHY) Specification for Low-Rate Wireless Personal Area Networks (LR-WPANs), Amendment 1: Add Alternate PHYs.

[20] USB Implementors Forum, Wireless USB 1.0 specification, http://usb.org, 2009.

[21] Anaxo, Ultra Wideband Communication Development Kit: http://www.ines.zhaw.ch/fileadmin/user_upload/engineering/_Institute_und_Zentren/INES/Wireless/UWB/Anaxo_-_Short_Form_Datasheet.pdf, June 2011

[22] INES - Institute of Embedded Systems : http://www.ines.zhaw.ch/en/engineering/ines/institute.html, June 2011

[23] Prytz, G.: A performance analyse of EtherCAT and PROFINET IRT, IEEE ETFA 2008

# Limits on the use of inertial actuators in active vibration control

**Simone CINQUEMANI, Ferruccio RESTA, Massimo MONGUZZI**

**Dipartimento di Meccanica, Politecnico di Milano,**
**Campus Bovisa Sud, via La Masa 34, 20156, Milano, Italy**

## ABSTRACT

Inertial actuators are widely used device in active vibration control applications. The aim of the paper is to analyze the limits on the use of such devices in terms of stability when fully active skyhook damping control is implemented. Different control strategies are investigated and numerical simulation are led considering different vibrating mechanical systems.

## 1. INTRODUCTION

Undesirable vibrations can induce noise, bad performance or even severe damage. Passive damping materials have been used effectively for a very long time, but especially for micro-amplitude and low frequency vibration, passive damping materials are inadequate and not effective. Hence active vibration control (AVC) techniques with feedback control feature have begun to be used to meet stringent accuracy and performance requirements. Direct velocity feedback control is one of the most commonly used strategy. It consists on a secondary force, which is designed to be proportional to the absolute structure velocity, which acts on the structure with the effect of increasing the damping of the structure itself. As it is known [7], the system is unconditionally stable for any positive real feedback gain.

A common way to generate the control force is by using inertial actuators. An inertial actuator consists on a suspended mass on a fixed base through an elastic element which is put in vibration by an internal force. This force can be generated through different physical principles, for example, electromagnetically. When the base is bound to the structure to be controlled, the force exerted by the actuator to the structure is equal to the force of inertia resulting from the acceleration of the suspended mass. Knowing the dynamics of the actuator, it is possible to control the exerted force acting on the internal force. The advantage of inertial actuators is they are mounted directly on the vibrating structure and do not need a fixed external base to react. On the other side, the dynamics of these actuators is strongly coupled with the one of the structure. Moreover, control techniques that are unconditionally stable in the case of ideal actuators (as the Skyhook logic), can become unstable when the force is generated by inertial actuators. For this reason, the dynamics of the actuators must necessarily be considered in the design of an active vibration controller.

Many controllers have been developed for the use of inertial actuators in AVC applications: an overview of linear controllers is described in [2]. Some nonlinear controls considering the saturation of the devices are described in [5], while a deep study on the stability can be found in [3], [6]. Benassi, Elliott and Gardonio have done several numerical and experimental studies on the stability and on the performance of inertial actuators used to control vibrations of a thin plate [9], [8], also considering the opportunity to use the feedback of the transmitted force [1] or the relative displacement of the inertial mass [4], in order to increase the stability margin.

The aim of this paper is to analyze the opportunity of using inertial actuators for the practical arrangement of a fully active skyhook control, highlighting its limits in terms of stability. After a brief description of the problem, the paper introduces the use of a compensator filter to delete the causes of instability in order to make the controlled system unconditionally stable.

The paper is structured as follows.
Section 2 recalls the basis of the functioning principle of inertial actuators. Section 3 shows the traditional approach in control design highlighting the limits on the use of inertial actuators in terms of system stability. The use of a compensating filter to consider the dynamic of the actuator is investigated showing the improvements obtained in terms of vibration reduction. Finally conclusions are drawn in Section 4.

## 2. FUNCTIONING PRINCIPLE OF INERTIAL ACTUATORS

An inertial actuator consists on a mass, free to move, connected to a fixed base through a spring-damper element. The base is connected to the vibrating structure and a force, whose origin can be different depending on the type of actuator, acts respectively on the inertial mass and on the structure itself. The type of actuators to be used depends on the maximum force to be transmitted and, above all, on the operating range of frequencies. Most commonly used actuators are electromagnetic shaker or magnetostrictive devices.

Considering the actuator as an inertial mass mounted on a vibrating structure through a spring-damper element

(Fig.1), its equation of motion is:

$$m_a \ddot{z}_a + r_a \left(\dot{z}_a - \dot{z}_e\right) + k_a \left(z_a - z_e\right) = -f_a \qquad (1)$$

where:

- $\dot{z}_a$ is the suspended mass velocity;
- $\dot{z}_e$ is the velocity of the vibrating structure;
- $f_a$ is the force generated by the actuator;
- $f_t$ is the force transmitted to the structure.

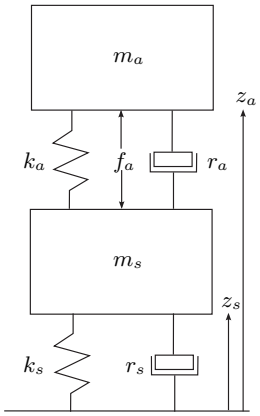



Figure 1: Mechanical scheme of an inertial actuator mounted on a single d.o.f. vibrating structure.

The force transmitted to the structure is equal to the force of inertia related to the actuator suspended mass:

$$f_t = T_a \, f_a - Z_{aa} \, \dot{z}_e = -m_a \ddot{z}_a \qquad (2)$$

where:

- $T_a = \dfrac{f_t}{f_a}$ is the actuator blocked response transfer function between the internal force generated by the actuator and the force transmitted to the structure, when the structure is still;
- $Z_{aa} = \dfrac{f_t}{\dot{z}_e}$ is the actuator mechanical impedance that is the transfer function between the structure velocity and the force transmitted when the actuator is not powered.

The two transfer functions can be obtained from the equations of motion (1). In the Laplace domain it results:

$$T_a(s) = \frac{F_t(s)}{F_a(s)} = \frac{m_a s^2}{m_a s^2 + r_a s + k_a} = \\ = \frac{s^2}{s^2 + 2h_a\omega_a s + \omega_a^2} \qquad (3)$$

where $h_a$ and $\omega_a$ are respectively the system adimensional dumping and natural frequency.

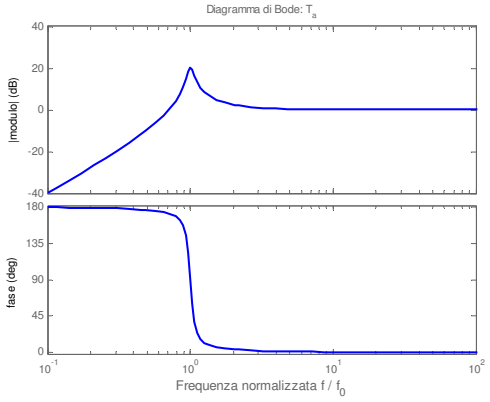



Figure 2: Actuator blocked response

Similarly, the mechanical impedance can be obtained from eq.(2), neglecting the term related to the internal force $f_a\,(s)$:

$$Z_{aa}(s) = \frac{F_t(s)}{v_e(s)} = m_a s \cdot \frac{r_a s + k_a}{m_a s^2 + r_a s + k_a} = \\ = m_a s \cdot \frac{2h_a\omega_a s + \omega_a^2}{s^2 + 2h_a\omega_a s + \omega_a^2} \qquad (4)$$

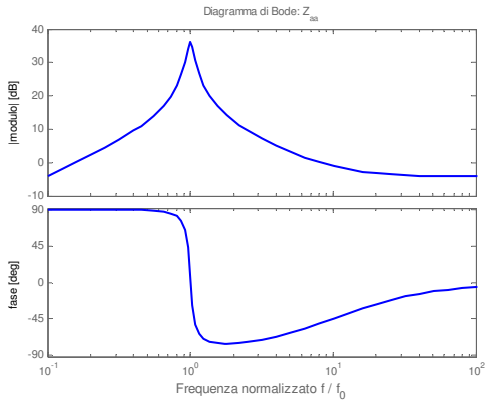



Figure 3: Actuator mechanical impedance

For frequencies higher than the actuator resonance, (Fig. 2, 3), both the two functions $T_a$, $Z_{aa}$ tend to have a constant magnitude and no phase shift. In this condition the actuator can be suitably considered as an ideal generator

of force. Vice versa, for lower frequencies, function $T_a$ shows a $+180^o$ phase shift, while its magnitude significantly changes with frequency. In the same range, the mechanical impedance has a $90^o$ phase shift.

## 3. DIRECT VELOCITY FEEDBACK CONTROL

In AVC applications, the amplitude of vibration of a structure van be reduced designing the control force to be proportional and in phase with the structure velocity and thus to increase the damping of the structure itself.

The skyhook control is always unconditionally stable if the actuator can be considered as ideal [7]. This condition is not verified when the control force is generated through an inertial actuator linked to the structure [3], [6], [9], [8].

This limit is essentially due to the hypothesis, done by many researchers. They are:

1. to damp vibrations only in structures whose natural frequencies are higher than the one of the actuator,
2. to neglect the dynamic of the actuator and to consider the internal force generated by the actuator equal to the force actually transmitted to the structure.

Due to these assumptions the system can not be unconditionally stable. In the following the limits of this approach are shown. The paper introduces a compensator filter that, considering the dynamics of the actuator, allows to design the control force in order to be proportional to the structure velocity and exactly in phase with it. The problem of stability is analyzed in detail, firstly considering a single degree of freedom structure and, subsequently, extending the results to $n$ d.o.f. vibrating systems.

### Single d.o.f. vibrating structure - Traditional approach

Consider a single d.o.f. vibrating structure whose natural frequency ($\omega_s$) is higher than the one ($\omega_0$) of the inertial actuator used to suppress vibrations.

Considering the system as shown in Fig.1, the equations of motion can be written as:

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_a \end{bmatrix} \begin{Bmatrix} \ddot{z}_1 \\ \ddot{z}_a \end{Bmatrix} + \begin{bmatrix} r_1 + r_a & -r_a \\ -r_a & r_a \end{bmatrix} \begin{Bmatrix} \dot{z}_1 \\ \dot{z}_a \end{Bmatrix} + + \begin{bmatrix} k_1 + k_a & -k_a \\ -k_a & k_a \end{bmatrix} \begin{Bmatrix} z_1 \\ z_a \end{Bmatrix} = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix} f_a \tag{5}$$

$$[M]\,\underline{\ddot{z}} + [R]\,\underline{\dot{z}} + [K]\,\underline{z} = [\Lambda_C]\,f_a \tag{6}$$

To increase the damping of the structure the control force is designed to be proportional and in phase with the structure absolute velocity:

$$f_a = -g_v \dot{z}_1 \tag{7}$$

Substituting eq. (7) in eq. (5) the effect of the control can be summarized considering the damping matrix of the controlled system.

$$([R] + [R_C]) = \begin{bmatrix} r_1 + r_a + g_v & -r_a \\ -r_a - g_v & r_a \end{bmatrix} \tag{8}$$

Since the resulting matrix is non symmetric, the controlled system has a stability limit as a function of the control gain $g_v$:

The frequency response of the system to a disturbing force is shown in Fig. 4. The solid line represents the behavior of the uncontrolled system, while the dashed line represents the behavior of the controlled one. It is noted that the control is able to effectively increase the damping of the structure, while it acts as a disturbance on the actuator inertial mass. The same conclusion can be drawn by observing the root locus of the coupled system.

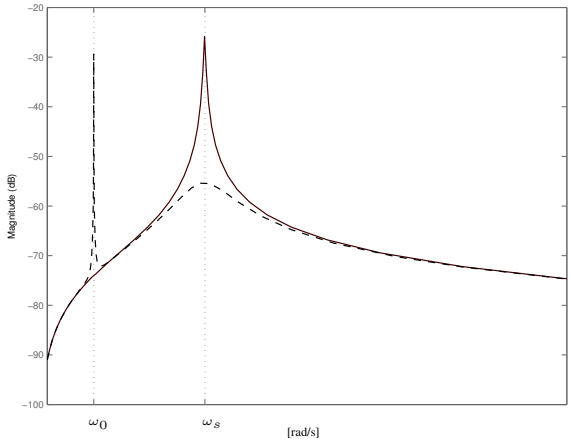



Figure 4: System frequency response: solid line - uncontrolled system; dashed line - controlled system

Same results can be achieved applying the criterion of Nyquist to the open loop transfer function:

$$L_1(s) = g_v \cdot \frac{\dot{z}_1}{f_a} = = g_v \cdot \frac{m_a s^3}{(s-p_1)(s+p_1)(s-p_2)(s+p_2)} \tag{9}$$

The shape of the open loop transfer function shows the controlled system is only conditionally stable. The Nyquist diagram presents a ring in the left-half plane (due to the dynamics of the inertial actuator), which tends to enclose the point (- 1, $0j$) when the gain of the control is increased.

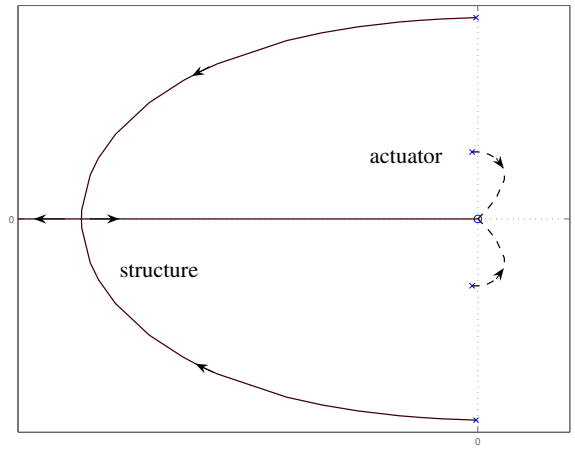



Figure 5: Root locus of the controlled system

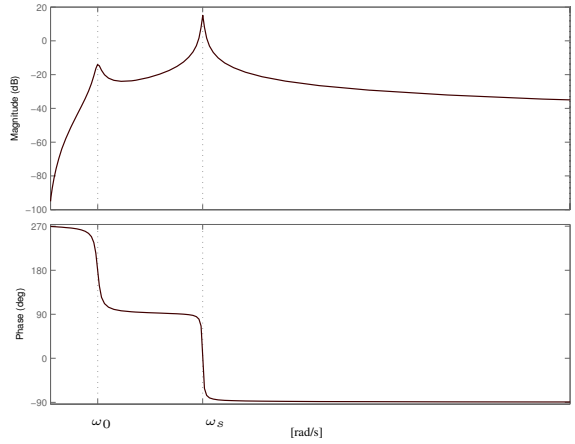



Figure 6: Open loop transfer function

If the vibrating structure were characterized by a natural frequency lower than the one of the inertial actuator ($\omega_s < \omega_0$) the problem remains unchanged. Since in the range of frequencies lower than $\omega_0$ the force transmitted has a $180^o$ phase shift with respect to internal force generated by the actuator, the equation (7) becomes:

$$f_a = +g_v \dot{z}_1 \tag{10}$$

**Single d.o.f. vibrating structure - Implementation of a compensator filter**

Limits on stability can be theoretically overcame by using a compensator filter that takes into account the dynamics of the actuator. The simplest filter can be designed as:

$$K(s) = T_a^{-1}(s) \tag{11}$$

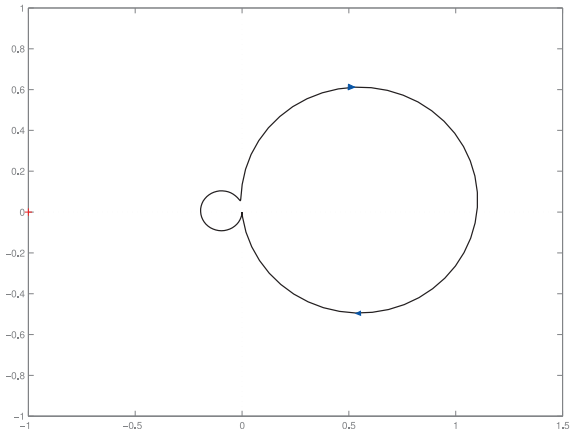


Figure 7: Nyquist plot of $L_1(s)$

Internal force generated by the actuator is:

$$f_a = -g_v \cdot \frac{1}{T_a}\dot{z}_1 = -k_p z_1 - k_d \dot{z}_1 - k_i \int z_1 dt \tag{12}$$

where:

$$\begin{cases} k_p = g_v \cdot 2h_a\omega_0 \\ k_d = g_v \\ k_i = g_v \cdot \omega_0^2 \end{cases} \tag{13}$$

To write the equation of motion of the controlled system in the time domain, the state of the variable integrated with respect to time has to be considered:

$$x_I = \int z_1 dt \rightarrow z_1 = \dot{x}_I \tag{14}$$

It results:

$$\left\{ \begin{array}{c} \underline{\dot{x}} \\ \dot{x}_I \end{array} \right\} = \left[ \begin{array}{cc} [A] & [0] \\ \left[\tilde{C}\right] & [0] \end{array} \right] \left\{ \begin{array}{c} \underline{x} \\ x_I \end{array} \right\} + \left[ \begin{array}{c} [B] \\ 0 \end{array} \right] f_a \tag{15}$$

$$f_a = -\underline{g}^T [C] \left\{ \begin{array}{c} \underline{x} \\ x_I \end{array} \right\} \tag{16}$$

where:

$$\left[\tilde{C}\right] = \left[ \begin{array}{cccc} 0 & 0 & 1 & 0 \end{array} \right]$$

$$[C] = \left[ \begin{array}{ccccc} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right] \tag{17}$$

are suitable matrix to extract corresponding quantities from the state and:

$$\underline{g}^T = \left[ \begin{array}{ccc} k_p & k_d & k_i \end{array} \right]$$

collects the gains of the controller.

The frequency response of the coupled system (actuator + structure) is shown in Fig. 8. The solid line represents the case of uncontrolled system, while the dashed line represents the behavior of the controlled system using the filter compensator described in (11) to cancel the dynamics of the actuator. It is noted that the control force is able to effectively increase the damping of the structure, without significantly changing the amplitude of oscillations of the vibrating mass.

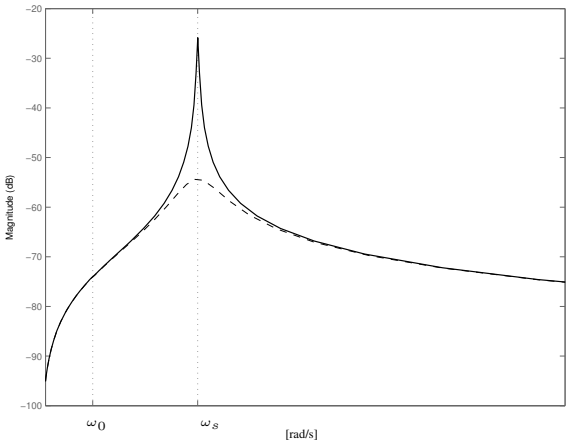



Figure 8: System frequency response: solid line - uncontrolled system; dashed line - controlled system

The stability of the controlled system can be assessed by the eigenvalues of the matrix of state equation (15), or by applying the Nyquist criterion to the open loop transfer function:

$$L_2(s) = \\ = \frac{g_v \cdot m_a s^3}{(s-p_1)(s+p_1)(s-p_2)(s+p_2)} \cdot \frac{(s-p_a)(s+p_a)}{s^2} \quad (18)$$

Differently to the previous case, the control results to be always stable. Analyzing the open loop transfer function it is evident that the zeros of the filter should compensate the poles associated with the inertial actuator resonance. At the same time the two poles in the origin require that the phase at zero frequency is +90$^o$. For this reason the Nyquist diagram has no loops in the left-half plane, ensuring the unconditional stability.

**Extension to $n$ d.o.f. structures**

Vibration control of a $n$ degrees of freedom structure summarizes the cases previously seen. The need to consider the dynamics of the suspended mass of the inertial actuator still remains, while adding the problem concerning the

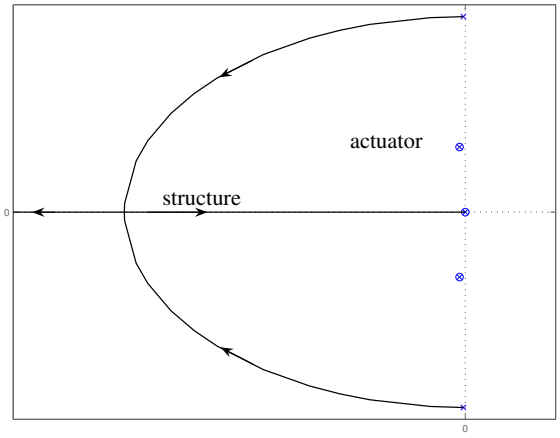



Figure 9: Root locus of the controlled system

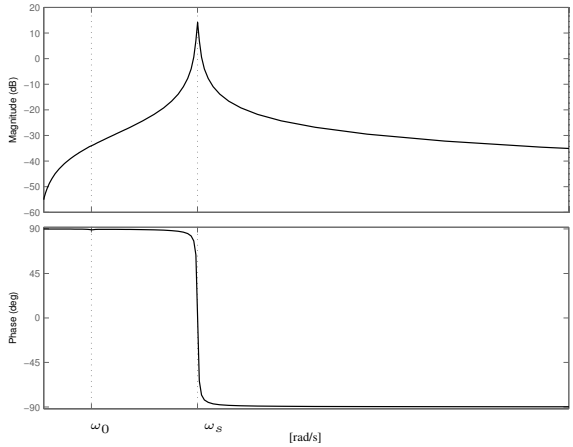



Figure 10: Open loop transfer function

possibility that the vibrating structure may have modes of vibration at frequencies even lower the resonance of the actuator.

Due to the transition of $180^o$ in the phase of the response blocked inertial actuator it is evident that it is not possible to have a control force in phase with the structure velocity simply designing the internal force to be proportional to it. In particular, while the traditional control increases the damping of structural modes whose frequencies are higher than the natural frequency of the actuator, at the same time it decreases the damping of structural modes with lower frequencies (Fig.12).

On the contrary, properly filtering the control signal using a compensator filter, the controlled system becomes unconditionally stable. All the structural modes can be effectively damped (Fig.13) and the oscillations of the suspended mass are not increasing.

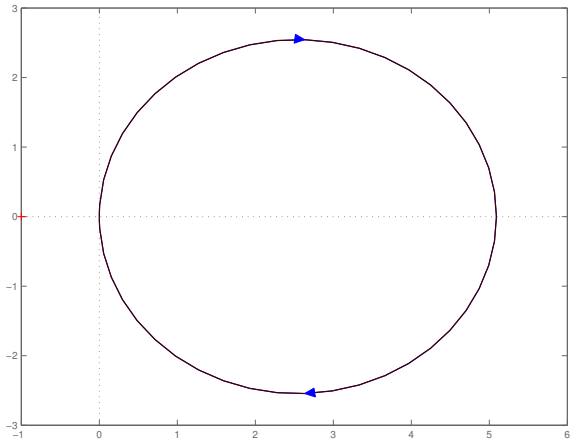


Figure 11: Nyquist plot of $L_2(s)$

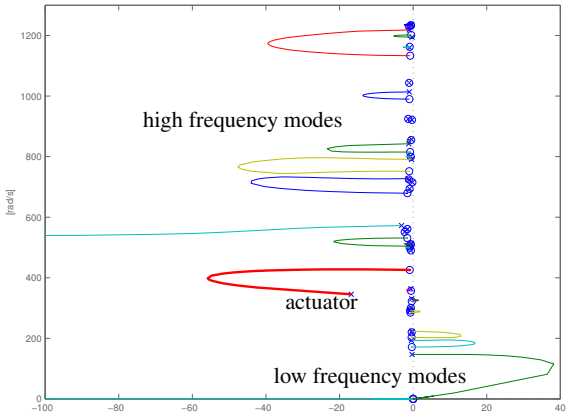



Figure 12: Root locus of a controlled $n$ d.o.f. structure with traditional approach

## 4. CONCLUSION

The use of inertial actuators in AVC application, using techniques known in the literature, implies a stability limit of the controlled system. It is due to the assumptions of neglecting the dynamics of the actuator considering it as ideal. The use of a compensator filter, describing the actuator dynamics, allows to overcome this limit and to make the controlled system unconditionally stable.

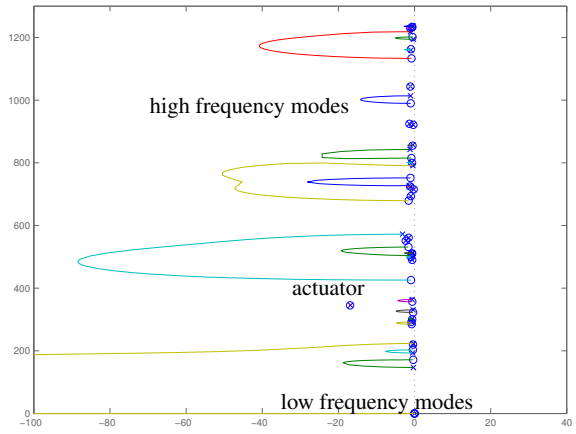



Figure 13: Root locus of a controlled $n$ d.o.f. using the compensator filter

## REFERENCES

[1] S. J. Elliott C.Paulitsch, P. Gardonio. Active vibration damping using an inertial, electrodynamic actuator. *Journal of Vibration and Acoustics*, 129:39–47, 2007.

[2] D. J. Inman D. C. Zimmerman. On the nature of the interaction between structures and proof-mass actuators. *Journal of Guidance*, 13:82–88, 1989.

[3] D. Borojevic D. K. Lindner, G. A. Zvonar. Nonlinear control of a proof-mass actuator. *AIAA Journal of Guidance, Control, and Dynamics*, 20:464–470, 1997.

[4] S. J. Elliott L. Benassi. Global control of a vibrating plate using a feedback-controlled inertial actuator. *Journal of Sound and Vibration*, 283:69–90, 2005.

[5] S.J. Elliott L. Benassi, P. Gardonio. Equipment isolation of a sdof system with an inertial actuator using feedback control strategie. part i: Theory. *ISVR Technical Memorandum*, 883, 2001.

[6] D. K. Lindner. Performance and control of proof-mass actuators accounting for stroke saturation. *AIAA Journal of Guidance, Dynamics, and Control*, 17, 1994.

[7] M.J.Balas. Direct velocity feedback control of large space structures. *Journal of Guidance*, 2(3), 1978.

[8] S. J. Elliott O. N. Baumann. Destabilization of velocity feedback controllers with stroke limited inertial actuators. *Journal of Acoustical Society of America*, 121:211–217, 2007.

[9] P. Gardonio S. J. Elliott, M. Serrand. Feedback stability limits for active isolation systems with reactive and inertial actuators. *Transactions of the ASME*, 123:250–261, 2001.

# Sliding mode controller for a 2 dof fully pneumatic parallel kinematic manipulator

**Hermes GIBERTI, Simone CINQUEMANI**

**Mechanical Engineering Department, Politecnico di Milano,**
**Campus Bovisa Sud, via La Masa 34, 20156, Milano, Italy**

## ABSTRACT

Sliding mode controller approach (SLM) was used in pick and place position control of a fully pneumatic parallel robot in this paper. The SLM approach could be used to overcome the nonlinearities associated with pneumatic systems and non-Cartesian kinematic. The paper shows the control law design procedure and includes simulation result to evaluate the performance achieved by the control algorithms in terms on positioning error. In order to prove the stability of the control law we tested it changing the payload moved by the robot end-effector.

## 1. INTRODUCTION

Mechanical industries are showing a growing interest to devices based on parallel kinematics, because these architectures often provide excellent performances in terms of stiffness/weight ratio if compared to traditional serial robots [8]. Another advantage of parallel architectures is that, in most cases, actuators can be placed on the truss, hence allowing the design of very light moving parts. This is a significant advantage, especially when powerful and huge motors are adopted. Indeed, the parallel kinematic arrangement of the links provides higher stiffness and lower moving masses that reduce inertia effects.

In many industrial applications, like the food and manufactory industry, very high precision positioning is not required. In these cases pneumatic actuators afford the opportunity to design a simple and low cost positioning system.

Pneumatic cylinders could be hardly used for serial robots, but they can be easily employed in a parallel device because cylinders and accessories would be simply fixed to the truss, thus achieving a limited weight for the moving parts. Moreover pneumatic cylinders may be directly coupled with a load without needing a transmission system. Unlike hydraulic actuators they are clean, reliable and they are inexpensive. Pneumatic cylinders have several advantages as high compactness, high acceleration capability, high robustness and reliability. Moreover pneumatic drive systems need low maintenance and thanks to the high standardization degree and to the ease of assembly they allow to realize modular systems.

Main limits are related to the difficulty of controlling the position and therefore the accuracy and repeatability of movements [2]: the pneumatic actuators have a non-linear behaviour due to the mechanical stick-slip friction and air compressibility that makes difficult and expensive the positioning control [11]. In fact, in the industrial fields the pneumatic actuators are mainly used in open-loop configuration. This allows fixed point positioning of the pay-load by means of mechanical end-stroke stops. Fully pneumatic systems with multi-degrees of freedom are therefore not used in industrial applications [9]. The characteristics of the parallel robots joined to the servo-pneumatic cylinders allow new solutions for industrial automation.

This paper has two main parts: in the first one we present the design and the mathematical model of a fully pneumatic 2 Dof robot. The robot is designed to carry out pick and place operations in industrial environments. This robot is a device based on parallel kinematic scheme [7] and it is built with 2 double acting rodless pneumatic cylinders placed in a planar frame. Pneumatic actuators are regulated by two 5/3 pneumatic proportional valves. The mathematical model of the pneumatic actuators dynamic and of the air flow through the inlet and outlet valves is based on elliptical approximation of non linear behaviour of pneumatic devices [1]. Friction and sticking phenomena are considered too.

The second part of the work relates the position control algorithm developed and the evaluation of its performances when applied to a multibody model of the manipulator. The control algorithm we used to control the piston position is a sliding mode control a kind of variable structure control [3]. Sliding mode control (SLM) has several benefits such as fast response and low sensitivity to disturbance and system parameter variations. SLM has been promoted as a solution to overcome the non-linearity associated with air compressibility in pneumatic positioning systems [10], [4]. One observe, the continuous form of sliding mode control (CLSM) does not require a mathematical model of the system to be controlled.

This feature makes the CSLM very useful to control complex and nonlinear systems like as that one we studied. Indeed the fully pneumatic parallel robot shows a non-Cartesian kinematic structures and a nonlinear actuation system. The paper shows the control law design procedure and it includes simulation result to evaluate the

Table 1: Nomenclature

| Symbol | Description |
|---|---|
| $q_1, q_2$ | Position of cylinder sliders |
| $l$ | Link length |
| $\alpha$ | Angular link position |
| $P_x, P_y$ | End-effector translating coordinates |
| $P_x^*, P_y^*$ | End-effector coordinates |
| $P_{ai}, P_{bi}$ | Cylinder chamber pressure |
| $G_{ai}, G_{bi}$ | Mass air flow |
| $P_0$ | Pressure in standard condition |
| $\rho_0$ | Air density in standard condition |
| $A$ | Cylinder useful section |
| $P_v$ | Downstream pressure |
| $P_m$ | Upstream pressure |
| $C$ | Sonic conductance |
| $\beta$ | Critical pressure ratio |

performance achieved by the CSLM control algorithms in terms on positioning error. In order to prove the stability of the control law we tested it changing the payload moved by the robot end-effector. Symbols used in the exposition are described in table 1.

## 2. ROBOT DESCRIPTION

A 2 dof robot has been designed to carried out pick and place operations for industrial applications. Figure 1 shows the kinematic model of the developed parallel kinematic manipulator. It consists on two rigid links, each one connected to a slider at one side and to the other link on the other side through revolute joints. Sliders are driven by pneumatic rodless actuators placed in a planar frame. The linear drive is a FESTO DGPL-32-1200-PPVA-B-KF-GK-SV commercial cylinder. The pneumatic positioning system is realized by a 5/3-way proportional directional flow control valve (FESTO, model MPYE-5-1/8-HF-010-b). The mechanical characteristic of the valves and cylinder are estimated in [5]. Such a manipulator is able to reach a point in the plane $xy$ inside the workspace. The orientation of the end effector is maintained thanks to an auxiliary link connected both to the end effector and one of the two sliders. Figure 1 shows the kinematic scheme of the manipulator under study. The end effector position ($P^*$) is described through its cartesian coordinates ($P_x^*$, $P_y^*$), while the displacements of the two pneumatic cylinders actuating the sliders are respectively $q_1$, $q_2$.

### Kinematic analysis

The kinematic analysis of the robot can be carried out considering the system as shown in Fig.1. Since the end effector ($P^*$) can only translate, its velocity and acceleration are the same of the ones related to the point $P(P_x, P_y)$. The end effector position can be expressed as a function of the actuator displacement as:

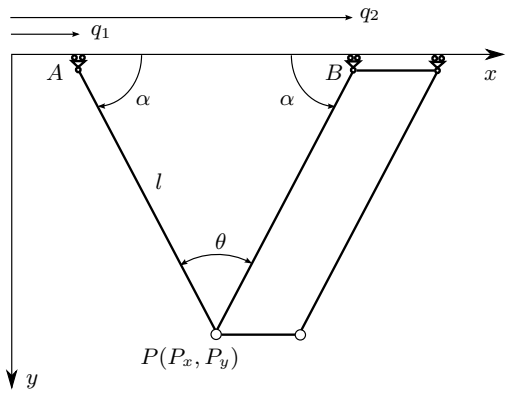



Figure 1: Kinematic robot scheme

$$\begin{cases} q_1 + le^{i\alpha} = P_x + iP_y \\ q_2 - le^{i\alpha} = P_x + iP_y \end{cases} \tag{1}$$

Projecting the first equation of (1) respectively on real and imaginary axis one gets:

$$\begin{cases} q_1 + l\cos\alpha = P_x \\ l\sin\alpha = P_y \end{cases} \tag{2}$$

and then:

$$\sin\alpha = \frac{P_y}{l}; \qquad \cos\alpha = \frac{1}{l}\sqrt{l^2 - P_y^2}$$

Equations (1) can be rewritten and finally it is possible to solve the inverse kinematic using equations (3):

$$\begin{cases} q_1 = P_x + \sqrt{l^2 - P_y^2} \\ q_2 = P_x - \sqrt{l^2 - P_y^2} \end{cases} \tag{3}$$

Deriving equations (3) with respect to time, one can obtain the values of the actuators velocities as a function of the end effector position ($P_x$, $P_y$) and velocity ($\dot{P}_x$, $\dot{P}_y$):

$$\begin{cases} \dot{q}_1 = \dot{P}_x + \dfrac{P_y\dot{P}_y}{\sqrt{l^2 - P_y^2}} \\ \dot{q}_2 = \dot{P}_x - \dfrac{P_y\dot{P}_y}{\sqrt{l^2 - P_y^2}} \end{cases} \tag{4}$$

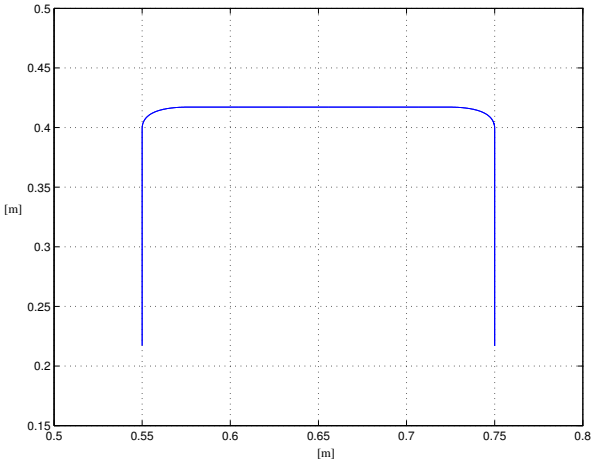



Figure 2: Trajectory in the $xy$ plane

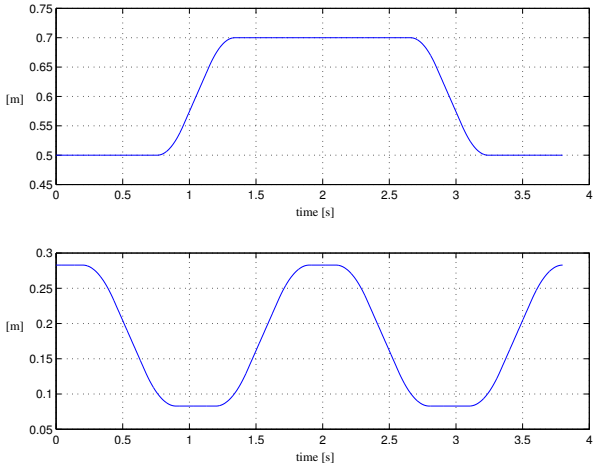



Figure 3: Motion law along the two axis $x$, $y$.

The same can be done for accelerations:

$$\begin{cases} \ddot{q}_1 = \ddot{P}_x + \dfrac{1}{\sqrt{l^2 - P_y^2}} \left[ \dot{P}_y^2 + P_y \ddot{P}_y + \dfrac{\left(P_y \dot{P}_y\right)^2}{\left(l^2 - P_y^2\right)} \right] \\ \ddot{q}_2 = \ddot{P}_x - \dfrac{1}{\sqrt{l^2 - P_y^2}} \left[ \dot{P}_y^2 + P_y \ddot{P}_y + \dfrac{\left(P_y \dot{P}_y\right)^2}{\left(l^2 - P_y^2\right)} \right] \end{cases} \quad (5)$$

Once the inverse kinematic problem is solved, it is possible to define a trajectory to be followed in the workspace (Fig.2) and the motion law to perform the movement along each axis. Figure 3 shows the designed law of motions along x, y axes. In this case constant symmetrical acceleration profiles have been chosen for both the axes.

Using equations (3), (4), (5) it is possible to calculate how to move the two sliders. These motion laws represent the reference of the controller needed to move the two pneumatic cylinders (Fig.4).

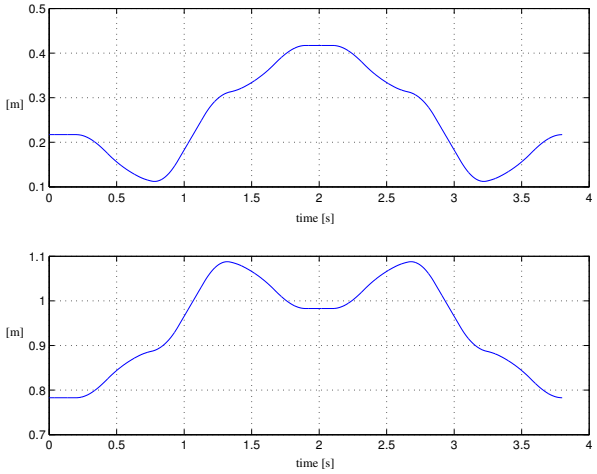



Figure 4: Motion law of each pneumatic actuator $q_1$, $q_2$.

## 3. DYNAMIC MODEL

The mathematical model of the system is composed by 3 main subsystems: the mechanical part describing the dynamic behavior of the manipulator, the pneumatic part concerning the dynamic of actuators and distribution circuitry and the controller that allows the end effector to reach the desired position.

### Mechanical model

The mechanical model is very simple and it is constituted by two sliders whose position, velocity and acceleration can be measured. Two forces, representing the action of the two pneumatic cylinders, are applied on the sliders. Such forces are calculated by the pneumatic subsystem. Rigid links are characterized by their inertial properties and are connected to the sliders as shown in Fig. 5. End effector position, velocity and acceleration is measured and it will be used by the controller to perform the desired movement.

### Pneumatics model

To complete the system study it is necessary to consider the equations of the pressure dynamics in the piston chambers and the equation of the mass air flow through the valves. With reference to equations [1], for every cylin-

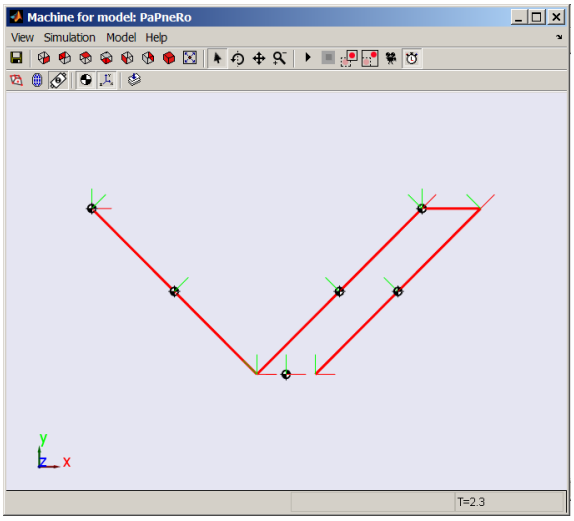



Figure 5: SimMechanics robot layout

der we have:

$$\begin{cases} \dot{P}_{ai} = \left[ G_{ai} - \rho_0 \left( \dfrac{P_{ai}}{P_0} \right) A\dot{d}_i \right] \dfrac{P_0}{\rho_0 A q_i} \\ \dot{P}_{bi} = \left[ G_{bi} + \rho_0 \left( \dfrac{P_{bi}}{P_0} \right) A\dot{d}_i \right] \dfrac{P_0}{\rho_0 A (c_i - q_i)} \end{cases} \tag{6}$$

where $P_{ai}$ and $P_{bi}$ are the pressures in the two chambers of *i*-th piston, $G_{ai}$ and $G_{bi}$ are the air flows entering in the two chambers by means of the valve, and $A$ is the useful section of the cylinder. The air flow according to the elliptic approximation can be expressed by the following relation:

$$G = \rho_0 C P_m \Phi \left( \frac{P_v}{P_m}, \beta \right) \tag{7}$$

were the function $\Phi$ is defined as:

$$\Phi = \begin{cases} 1 & \text{for } \dfrac{P_v}{P_m} \leq \beta \\ \sqrt{1 - \left[ \dfrac{\frac{P_v}{P_m} - \beta}{1 - \beta} \right]^2} & \beta < \dfrac{P_v}{P_m} \leq 1 \end{cases} \tag{8}$$

where $C$ and $\beta$ are respectively the conductance and the critical pressure ratio (these are empiricist parameters which define the behavior of valves; in order to their characterization it is possible to use the following regulation [1]) while $P_m$ and $P_v$ are the upstream and downstream pressure regarding to the air flow direction.

## 4. Continuous Sliding Mode Controller

The implemented control scheme is based on the variable structure control theory, in particular on sliding mode control. It is easy to demonstrate that servo-pneumatic system is a third-order servosystem. In fact the dynamic equation of a symmetric pneumatic cylinder can be written as:

$$M\ddot{x} + \gamma\dot{x} + f = A(P_a - P_b) \tag{9}$$

where $M$ is the load mass, $\gamma$ is a friction coefficient, $x$ is a piston position, $P_{a,b}$ are the chamber pressures and $f$ is an external force applied to the system. Taking differentiation on both sides we get:

$$M\dddot{x} + \gamma\ddot{x} + \dot{f} = A(\dot{P}_a - \dot{P}_b) \tag{10}$$

Substituting the derivative of chamber pressure 6 in equation 10 and equation of mass air flow 7 in equation 6, we could introduce the control input in servosystem dynamic. Indeed the conductance $C$, shown in equation 7, depends on valve command signal. We have shown that the servo-pneumatic system is a third-order, but the precise knowledge of term $\dot{f}$, $\gamma$ is normally unavailable furthermore the approximations introduced on chamber pressure dynamics make inaccurate the mathematical description of the system. Luckily, one of the attractions of a CSLM is intended to be the ability to design a controller without the perfect system knowledge. This feature is very useful because we can design the controllers for every cylinder in a decoupled manner. The CSLM controller is based on the switching function $s$ expressed by:

$$s = m_1 e + m_2 \dot{e} + \ddot{e} \tag{11}$$

where $m_1$ and $m_2$ are the coefficients that define the sliding surface, as the terms $e$, $\dot{e}$ and $\ddot{e}$ are respectively the position, velocity and acceleration errors. The control law for CSLM can be given [3] as:

$$u = K \cdot X + V \cdot \text{sign}(s) \tag{12}$$

with $K$ and $X$ as gain and state vectors respectively; $V$ is the maximum supply voltage to the proportional valve. The first term $(K \cdot X)$ is often ignored in the applications where the problem is not reaching the sliding surface but staying on the sliding surface. The second term, causing the function sign$(s)$, may generate an excessive chattering in $u$ once the sliding surface is reached. It is important to reduce chattering phenomena when one uses a 5/3 way proportional pneumatic valve. One common technique to do this is to modify the basic control law of equation 12 to include a *boundary layer* [6]:

$$u = V \cdot \text{sat}(s) \tag{13}$$

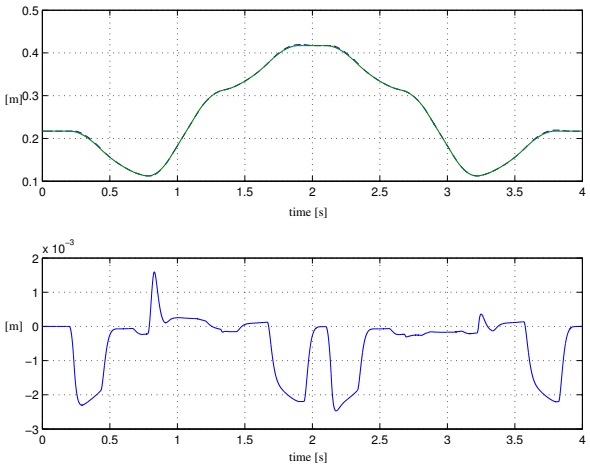



Figure 6: Comparison between the reference and the actual position $q_1$ (up) and the error between them (down).

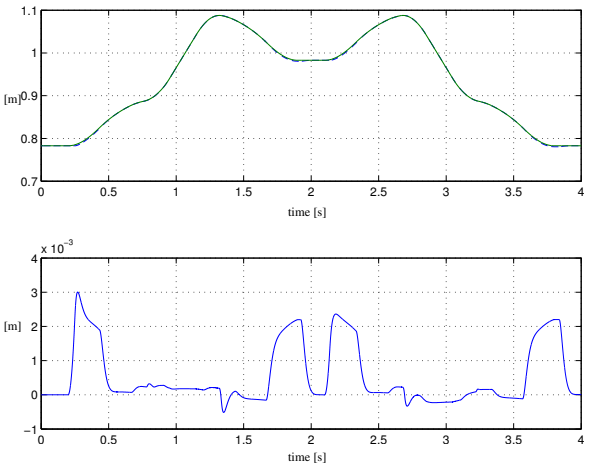



Figure 7: Comparison between the reference and the actual position $q_2$ (up) and the error between them (down).

where

$$\text{sat}(s) \quad = \quad \begin{cases} \text{sign}(s/\phi) & \text{when} \quad |s| > \phi \\ s/\phi & \text{when} \quad |s| \leq \phi \end{cases}$$

where $\phi$ is the parameter to define the boundary layer. It must be found to provide the proper balance between minimal chattering and acceptable accuracy.

## 5. RESULTS

The control action is tested on pick and place task shown in figures 2, 3 and 4. The payload is varied four time: 0.5 [kg], 2.5 [kg], 5.0 [kg] and 10.0 [kg] respectively. The controller parameters are tuned considering the small load and they remain the same during the different payload tests. In particular the sliding surface was defined by $m_1 = 2304$, $m_2 = 80$ moreover the boundary layer parameter was $\phi = 1.5$. The performance of the manipulator can be analyzed comparing the actual displacement of the two actuators with their reference values. Figures 6, 7 show the comparison between the reference and the actual position (up) and the error between them (down) respectively for joint coordinates $q_1$ and $q_2$ when the payload is 0.5 [kg]. The maximum position error is about 3mm, it is important to highlight that such error is limited to the motion phase, where it is not so important to exactly follow the desired trajectory. Errors in points where pick and place operations are carried out are very small.

Figure 8 shows the air pressures inside the two chambers of each cylinder while figure 9 shown the supply voltage to the two servovalve (in this case the maximum value is 5 [V]).

We can observe the smoothing effect due to boundary layer action. Figure 10 shows the effect of changing payload value on CLSM performance in terms of position error and supply voltage to the servovalve for one of sevopneumatic system in the first 1.5 [s] of the motion. The controller is robust when the mass is varied from 0.5 [kg] to 10 [kg].

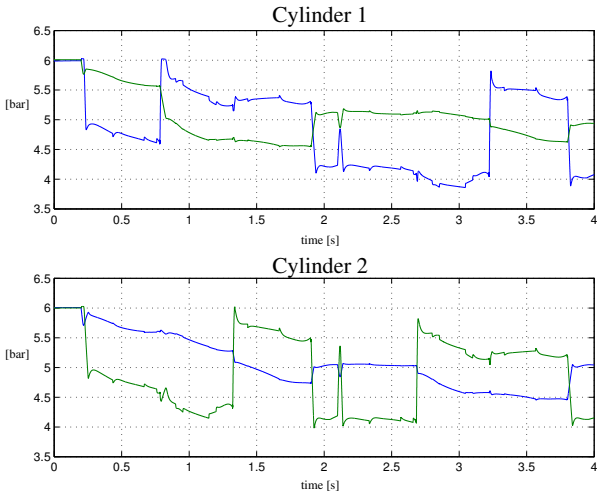



Figure 8: Air pressures inside the two chambers of each cylinder.

## 6. CONCLUSION

Pneumatic technology in parallel kinematics structures allows to develop new solutions to be proposed in industrial field which are characterized by ease of implementation, low cost and performance appropriate to the requirements of the industry. The CSLM controller is able to maintain

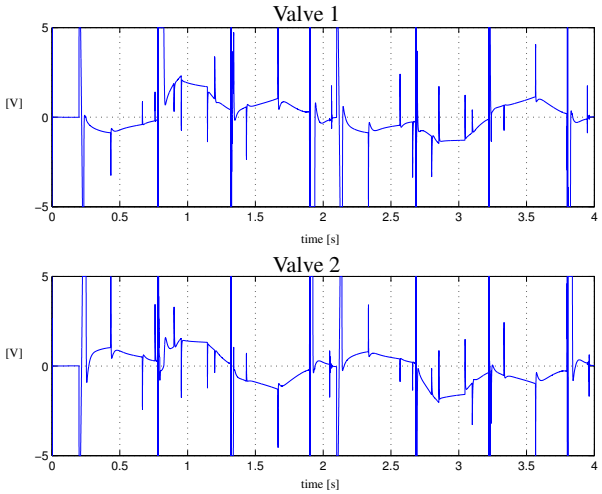



Figure 9: Supply voltage to the two servovalve.

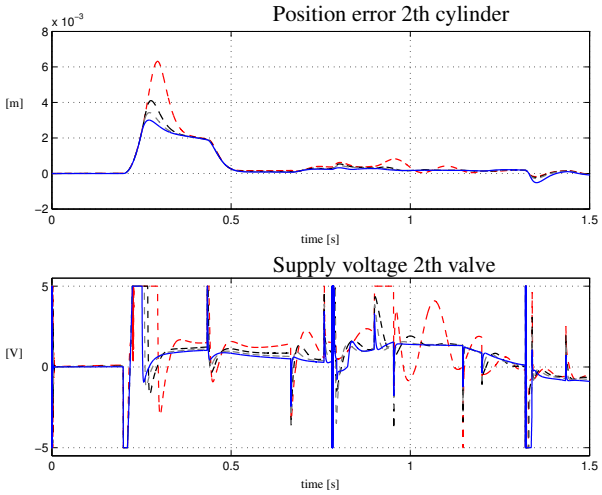



Figure 10: Effect of changing payload value on CLSM performance.

performance when the load mass is varied. This is a very good feature in a pick and place task because normally in the first part of the task, the robot ed effector moves with a full load while, in the second part, it returns empty.

Though the sliding mode control approach does not need an explicit mathematical model of the system only the knowledge of the system dynamics is required to select better sliding surface. Furthermore the control implementation is very easy and does not require an expensive hardware. The good result presented here suggest to further investigate the use of this technology (Parallel pneumatic robot - Continuous sliding mode controller) in industrial applications.

## REFERENCES

[1] Pneumatic fluid power-components using compressible fluids determination of flow-rate characteristics, iso 6358, 1989.

[2] X. Brun, M. Belgharbi, S. Sesmat, D. Thomasset, and S. Scavarda. Control of an electropneumatic actuator: Comparison between some linear and nonlinear control laws. *Proceedings of the Institution of Mechanical Engineers.Part I: Journal of Systems and Control Engineering*, 213(5):387–406, 1999.

[3] C. Edwards and S.K. Spurgeon. *Sliding mode control: theory and applications - ISBN: 0-7484-0601-8.* Taylor and Francis, London, United Kingdom, 1998.

[4] H. Giberti and P. Righettini. A nonlinear controller for trajectory tracking of pneumatic cylinders. *7th International Workshop on Advanced Motion Control*, pages 396–401, 2002.

[5] H.Giberti, P.Righettini, S.Chatterton, and R.Strada. Experimental setup and simulations of a parallel pneumatic robot. In *7th International workshop on research and education in mechatronics*, Stockholm, Sweden, 16-17 June 2006.

[6] Slotine J.J.E. Sliding controller design for nonlinear systems. *International Journal of Control*, 40(2):421–434, 1984.

[7] X.J. Liu, J. Wang, and G. Pritschow. On the optimal kinematic design of the prrrp 2dof parallel mechanism. *Mechanism and Machine Theory*, 41:1111–1130, 2006.

[8] J.P. Merlet. *Parallel Robots - ISBN: 14-02-04132-2*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.

[9] R.E. Rubio, S.L. Hernandez, S.R. Aracil, P.R. Saltaren, and J.A. Guerra. Implementation of decoupled model-based controller in a 2-dof pneumatic platform used in low-cost driving simulators. *CERMA 2009 - Electronics Robotics and Automotive Mechanics Conference*, pages 338–343, 2009.

[10] B.W. Surgenor and N.D. Vaughan. Continuous sliding mode control of a pneumatic actuator. *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME*, 119(3):578–581, 1997.

[11] F Xiang and J Wikander. Block-oriented approximate feedback linearization for control of pneumatic actuator system. *Control Engineering Practice*, 12(4):387–399, 2004.

# Evaluation for the Power Conscious Optimum Design of a Ubiquitous Processor

**Masa-aki FUKASE[*], Naomichi MIMURA[*], Kazuki NARITA[*], Tatsuya TAKAKI[*], Harunobu UCHIUMI[*], Takumi ISHIHARA[*], and Tomoaki SATO[†]**

***Graduate School of Science and Technology**

**†C&C Systems Center**

**Hirosaki University**

**Hirosaki, 036-8561, Japan**

**ABSTRACT**

In order to seek a leading edge processor of next generation information and communication technologies, the authors have so far developed a ubiquitous processor HCgorilla. It exploits the overall status of ubiquitous computing that are power consciousness for mobility, high performance for multimedia processing, and reliability for world wide divergence. The basic organization of HCgorilla follows multicore for bidirectional communication and multiple pipelines for ubiquitous computing. These are Java compatible media pipes and cipher pipes with practical cipher strength built in hardware RNG (random number generators). In order to fully utilize those architectural aspects described above, the role of semiconductor process is very important. Thus, this paper evaluates the potential of an HCgorilla chip by using a 0.18-μm CMOS standard cell process. The evaluation covers power consumption, throughput in both GIPS and GOPS, and cipher strength. With respect to these viewpoints, it is very important to suppress the power consumption of HCgorilla. In order to achieve useful knowledge for power conscious design of HCgorilla, power consumption is further analyzed more in detail. Then, the effect of gated clocking for power saving has been made clear.

**Keywords:** Ubiquitous processor, chip, power consciousness, wave-pipeline, throughput.

## 1. INTRODUCTION

Ad-hoc network is really an emerging technology for the next generation ubiquitous computing. It is contributive to the development of ubiquitous environment in view of cost performance, simplicity, functionality, usability, etc. However, ubiquitous community has left fundamental issues unsolved. Some of them are digital divide, information flood, etc. For these issues, software approach such as dependable computing is usually taken. But it is really time and power consuming in computing huge amount of multimedia data. Actually, massive data is used for the expression of multimedia information. This is crucial for the interaction between ubiquitous devices and human being. The massive quantity of multimedia information is very difficult for regular techniques like embedded software to satisfy various demands for not only security but also usability, speed, and power consciousness.

More fundamental issue is notorious security threat due to the expansion or diversity of ubiquitous platforms. Although the diversity is inevitable for the usability and functionality of ubiquitous network, it also causes notorious security issues like insecurity, security threat or illegal attack such as tapping, intrusion, pretension. The worldwide diversity vs. security threat is the two-faced characteristics of ubiquitous network. In order to keep the security of such a transient network, a practical solution is not to develop an extremely strong cipher scheme, but to explore a temporary security with practically enough cipher strength and without relying on permanent network infrastructure. Not only transience but also secureness is the one of most characteristic features of ad-hoc network.

An overall strategy for contribution to ad-hoc network will be power conscious hardware approach that develops a sophisticated single VLSI chip processor, which treats multimedia date with practical security over ubiquitous network. It is really worthwhile to exploit the hardware integration of possible key technologies.

Accordingly, the authors have so far developed HCgorilla [1]. HCgorilla is called a ubiquitous processor because it exploits the overall status of ubiquitous computing that are power consciousness for mobility, high performance for multimedia processing, and reliability for world wide divergence. The basic organization of HCgorilla follows multicore for bidirectional communication and multiple pipelines for ubiquitous computing. These are Java compatible media pipes and cipher pipes with practical cipher strength built in hardware RNG (random number generators).

This work is supported by VLSI Design and Education Center (VDEC), the University of Tokyo in collaboration with Synopsys, Inc. and Cadence Design Systems, Inc.

In order to fully utilize those architectural aspects described above, the role of semiconductor process is very important. Thus, this paper evaluates the potential of an HCgorilla chip by using a 0.18-μm CMOS standard cell process. The evaluation covers power consumption, throughput in both GIPS and GOPS, and cipher strength. With respect to these viewpoints, it is very important to suppress the power consumption of HCgorilla. In order to achieve useful knowledge for power conscious design of HCgorilla, power consumption is further analyzed more in detail. Then, the effect of gated clocking for power saving has been made clear.

## 2. PROCESSOR STRUCTURE

Fig. 1 shows the basic structure of the latest version of HCgorilla chips. Each of two symmetric cores is composed of two media pipes and a cipher pipe. The media pipe is a sort of an interpreter type Java CPU. The media instructions are free from scheduling because the media pipe's execution stage is made by a waved MFU. Table 1 summarizes the design environment of HCgorilla chips.

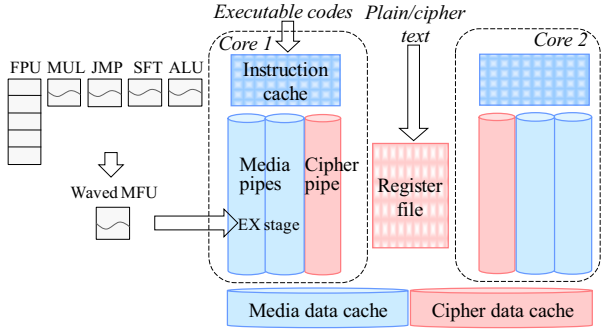



Fig. 1. Basic structure of HCgorilla.

Table 1. Design environment.

| | |
|---|---|
| ***Hardware*** | |
| *Work station* | A Pro OPT254 W/NVS285 RHL |
| *CPU* | AMD Opteron 3.0GHz |
| *Main memory* | 6144MB |
| *Swap memory* | 1024MB |
| ***Software*** | |
| *OS* | Red Hat Linux 4/SentOS 5.4 |
| *Synthesis tool* | Synopsys - Design Compiler D-2010.03 |
| *Simulation tool* | Synopsys - VCS version Y-2006.06-SP2 |
| ***Language*** | |
| *Synthesis* | VHDL |
| *Simulation* | Verilog - HDL |
| ***Technology*** | |
| ROHM 0.18μm CMOS Standard Cell Library | |

## 3. EVALUATION

Table 2 summarizes chip parameters and specifications of HCgorilla chips. The latest version, HCgorilla.6 is evaluated by using the design environment shown in Table 1. Clock frequency is derived from timing analysis. Other features are evaluated by using a test program shown in Fig. 2. This is used for the evaluation of both the cipher pipe and media pipes. The overall internal behavior in running the test program is also illustrated. The media pipe adds from 1 to 128. The difference of the routines A, B, and C is that the variable *k* and loop count are integers or floating point numbers. These routines are divided into four threads and assigned into four arithmetic media pipes in order to fully make use of the parallelism of arithmetic EX stage.

Table 2. Specifications of HCgorilla chips.

| Version | | | | HCgorilla.3 | HCgorilla.4 | HCgorilla.5 | HCgorilla.6 |
|---|---|---|---|---|---|---|---|
| Chip parameter | Design Rule | | | ROHM 0.18μm CMOS | | | |
| | Area | Chip | | 5.0mm × 7.5mm | | | |
| | | Core | | 4.28mm × 6.94mm | | | |
| | Power Supply | | | 1.8V (I/O 3.3V) | | | |
| Per processor | Power consumption | | | 241mW | | 274mW | |
| | Clock frequency | | | 330MHz | 400MHz | 200MHz | |
| | Instruction cache | | | 16bit × 32word × 2 | | 16bit × 64word × 2 | |
| | Data cache | | | 16bit × 128word | | 16bit × 128word × 2 | |
| | Stack memory | | | 16bit × 8word × 4 | | 16bit × 16word × 8 | |
| | Register file | | | 16bit × 64word | | 16bit × 128word | |
| | No. of cores | | | 2 | | | |
| Pipelines/ core | Arithmetic | No. | | 2/core | | | |
| | | Per arithmetic pipe | EX IU | 2-wave | | 2-wave MFU | |
| | | | EX FPU | NA | 5-clock | | |
| | | | Stacks | 1 | | 2 | |
| | | ILP degree | | 2 | | 4 | |
| | | Throughput | | | 0.07-0.09 GIPS | 0.16 GIPS | |
| | | Java bytecode | | 61 | 77 | 58 (102) | |
| | Cipher | No. | | 1/core | | | |
| | | RNG | | 4-bit LFSR × 1 | | 6-bit LFSR × 1 | 6-bit LFSR × 2 |
| | | Throughput | | 0.1-0.2 GOPS | | | |
| | | SIMD cipher code | | 2 | | | |

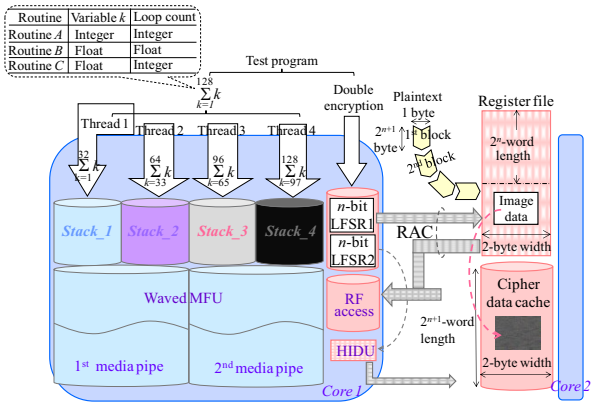



Fig. 2. Test program and internal behavior.

The cipher pipe does double encryption during the transfer of an image data from the register file to data cache. Due to restricted hardware quantity, the built-in RNG is a LFSR or linear feedback shift register. Register file and data cache is word structured. Each word is 2-byte width. So, a pixel occupies one and a half word. Register file stores a block as

follows. Plain and cipher texts are assumed to be divided into blocks as is similar to AES (Advanced Encryption Standard) and DES (Data Encryption Standard). Since the block size influence the performance and the secureness of cryptography, further expanding the register file and data cache is crucial. While LFSR1 controls the transposition cipher called RAC (random number addressing cryptography), LFSR2 controls a substitution cipher implemented by the hidable unit, HIDU.

### 3.1 Power

Fig. 3 shows the register file length dependency of HCgorilla.6's power consumption. Since it is derived from static analysis, occupied area is also shown.

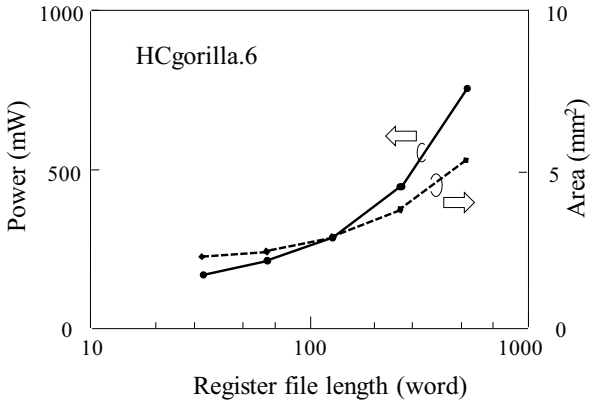



Fig. 3. Power consumption vs. register file length.

### 3.2 Throughput

Fig. 4 shows the result of evaluation for media processing performance in GIPS (giga instructions per second). Routines A, B, and C are described in Fig. 2.

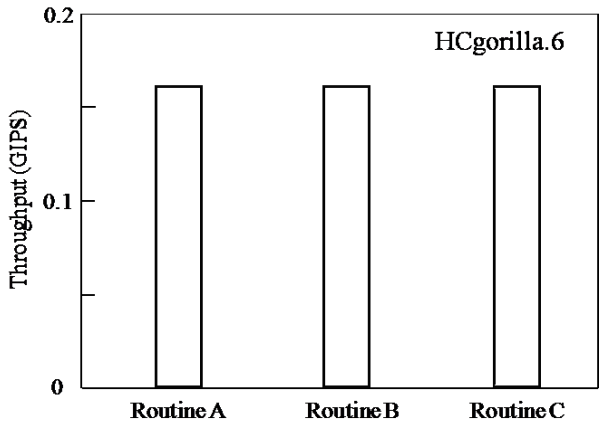



Fig. 4. Media pipe's throughput.

Fig. 5 shows HCgorilla.6's encryption/decryption performance in GOPS (giga operations per second) vs. register file length. The interpolation of an image data is the target of encryption and decryption. The time to be taken by rewriting register file is referred to DMA mode transfer. The throughput almost saturates at 128-word length.

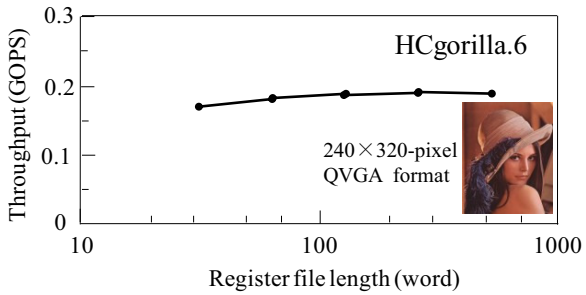



Fig. 5. Cipher pipe's throughput vs. register file length.

### 3.3 Cipher strength

As shown in Fig. 6, the cipher strength is measured in a ubiquitous environment where HCgorilla built-in ubiquitous platforms are used. In view of computational complexity, the strength or the degree of enduringness is evaluated by the time taken for the attack and the discovery of a true key. While decryption is the right recipient's process to recover the plaintext by using the given key, decipher, break, or crack is the third party's action to seek out an unknown key.

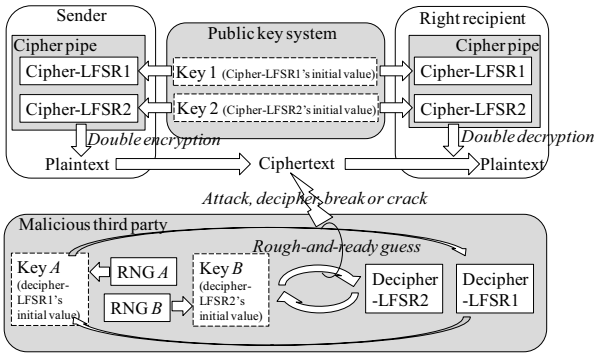



Fig. 6. Measurement of cipher strength.

Fig. 7 shows the register file length dependency of the cipher strength. As is illustrated in Fig. 2, it is a critical factor of cipher strength. From Fig. 7, the register file length is directly concerned with the cipher strength of HCgorilla.

## 4. DISCUSSION

The reason why various aspects of register file length dependency are evaluated is because the size of available memory space is one of critical factors of evry processors.

Especially, register file length is very important in designing cipher pipe. It also strongly affects power consumption of HCgorilla.

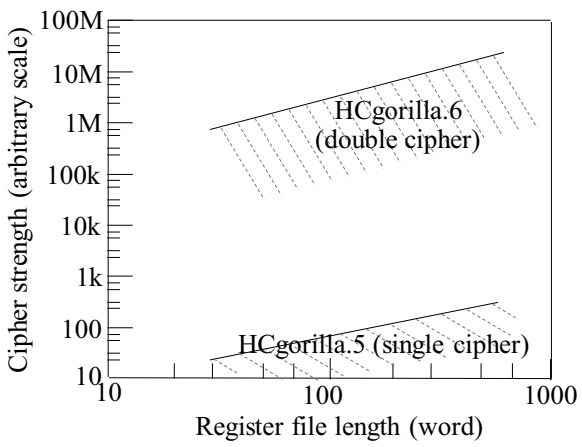



Fig. 7. Cipher strength vs. register file length.

In view of throughput, the dependency on register file length is weak from Fig. 5. Then, longer length is more desirable for cipher strength from Fig. 7. However, it is not so practical because power consumption rapidly increases when register file length is larger than 128 words as shown in Fig. 3. Thus, it is very important for the application of HCgorilla in ubiquitous network is to suppress the power consumption of HCgorilla.

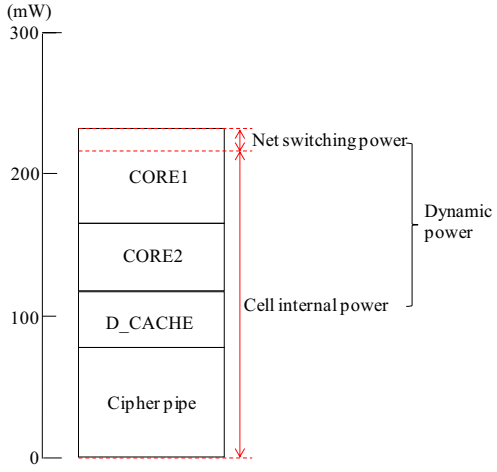



Fig. 8. Details of HCgorilla's power consumption.

In order to achieve useful knowledge for power conscious design of HCgorilla, power consumption is analyzed more precisely from dynamic analysis in running the routine A shown in Fig. 2. Fig. 8 shows the details of power consumption of various parts of HCgorilla. Here, the routine A is executed by the core 1. The components of power consumption are defined as follows.

Dynamic power
= cell internal power + net switching power (1)
Cell internal power =power consumption within cells
= short circuit power + switching power (2)
Net switching power
=power consumed by stray capacitance
dangling from the external loads (3)

It is clear from Fig. 8 that most of the power is exhausted dynamically and each part consumes similar power regardless of the contribution in running the test program. The core 2 is idle in this case. Data cache is accessed in very short period when the core 1 finishes the execution of the routine A. According to the evaluation of the power dissipation shown in Fig. 8, suppressing the dynamic power is crucial for power saving.

However, it is rather complicated to suppress dynamic power. Since Eqs. (1)-(3) are general and rather vague, let us consider a guide to this problem by using the static evaluation in the following [2].

$$P = \sum_{all\_cell} P_{cell} = \sum_{the\_clocked\_cell} P_{cell} \quad (4)$$

$$P_{cell}=A\{CV^2f+\tau VI_{short}f\}+VI_{leak} \quad (5)$$

$$f_{max}=(V-V_{th})^2/V \quad (6)$$

Here,

P: power consumption of a processor

$P_{cell}$: power consumption of a cell

A: switching probability of a clocked active cell within a clock cycle

$C=C_{cell}+C_{load}$

$C_{cell}$: stray capacitance dangling internally within the cell

$C_{load}$: stray capacitance dangling from the external loads

V: power supply voltage

$V_{th}$: threshold voltage of a MOS transistor

f: clock frequency

$f_{max}$: maximum clock frequency

τ: transient time during which short current flows from power supply to grand via active cell at on-off conversion

$I_{short}$: average of short current during τ

$I_{leak}$: leak current.

If V is made to decrease to suppress the dynamic power shown in the first and second terms of Eq. (5), $f_{max}$ also decreases from Eq. (6). Certainly, this deteriorates throughput in both GIPS and GOPS. Since cell-based approach for power saving based on Eqs. (4) and (5) accompanies complicated tradeoff, the alternative solution based on microarchitecture level will be

more preferable.

Gated clocking is a well known technology for microarchitecture level power saving. It stops the clocking of such circuit blocks with low activity that waste switching power. Since leakage power is extremely small in the case of the 0.18μm CMOS standard cell process used in this study, the gated clock is very effective for power saving. However, it is not directly available for HCgorilla due to following reasons. While CAD tools supporting gated clocking are implicitly specialized for regular pipelines, HCgorilla is wave-pipelined [3].

Taking into account of the manual tuning for the wave-pipelining of an HCgorilla's execution stage, another manual approach for the gated clocking in conjunction with the wave-pipelining is studied. In view of switching probability, the stack access stage and the waved MFU stages are the target of gated clocking as shown in Fig. 9. In addition, scan logic for DFT (design for testability) is introduced. Since the scan logic is applied to pipeline registers, a clock scheme to merge gated clocking and scan logic is developed in this study.

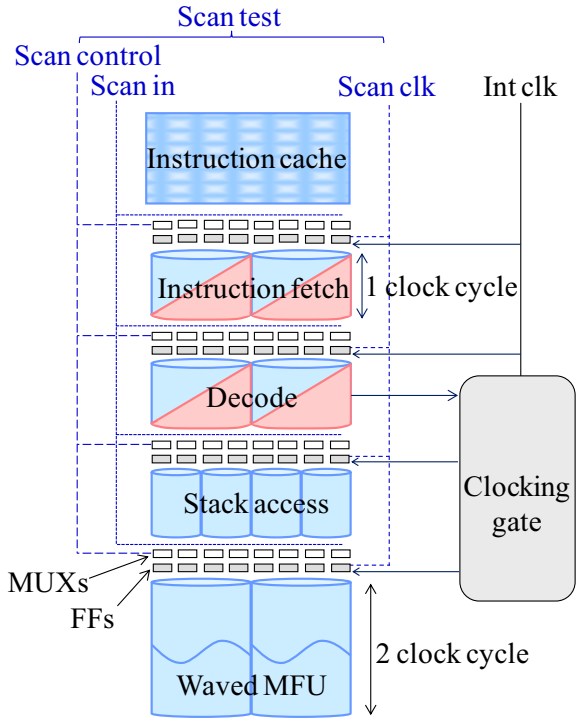



Fig. 9. Clock scheme for the combination of wave-pipelining, gated clocking, and scan logic.

The clock scheme shown in Fig. 9 is implemented in the context as follows. The pipeline registers are made to play the role of a shift register. The role is determined by a scan control signal. The serial mode of the shift register is used in the validation of the HCgorilla chip. On the other hand, a clocking gate unit is placed after an instruction decode stage produces a gated clock.

The table 3 shows the effect of gated clocking in running a simple integer summation program. Since HCgorilla's media pipe is Java compatible and thus follows a stack machine, it wastes large power in the stack access stage. Thus, the application of gated clocking in the stack access stage is very effective.

Table 3. Effect of gated clocking.

| | Stack access | Waved MFU |
|---|---|---|
| Nongated clocking | 22mW | 0.91mW |
| Gated clocking | 0.34mW | 0.11mW |

## 5. CONCLUSION

With the aim of applying HCgorilla in ubiquitous network, the latest version has been evaluated in view of power consumption, throughput in both GIPS and GOPS, cipher strength. As a result, it is very important to suppress the power consumption of HCgorilla. In order to achieve useful knowledge for power conscious design of HCgorilla, power consumption is analyzed. Then, the effect of gated clocking has been made clear. The next step of this study is to implement the clock scheme for the combination of wave-pipelining, gated clocking, and scan logic in the real chip of HCgorilla.

## 6. REFERENCES

[1] M. Fukase and T. Sato, "H/S Collaborative Development of a Ubiquitous Processor Free from Instruction Scheduling and Pipeline Disturbance," Proc. of ICIS 2010, pp. 57-62, 2010.

[2] T. Mudge, "Power: A First-Class Architectural Design Constraint," Computer Magazine, Vol. 34, No. 4, pp. 52-58, April, 2001.

[3] M. Fukase, A.Yokoyama, T. Ishihara, H. Uchiumi, and T. Sato, "Wave Degree versus Dominant Characteristics of a Waved Multifunctional Unit," Proc. of CCCT 2010, Vol. II, pp. 164-168, 2010.

# Delay Time Analysis of Reconfigurable Firewall Unit

**Tomoaki SATO**
**C&C Systems Center, Hirosaki University**
**Hirosaki 036-8561 Japan**

**Phichet MOUNGNOUL**
**Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang**
**Bangkok 10520 Thailand**

**Masa-aki FUKASE**
**Graduate School of Science and Technology, Hirosaki University**
**Hirosaki 036-8561 Japan**

## ABSTRACT

A firewall function is indispensable for mobile devices and it demands low-power operations. To realize this demand, the authors have developed a firewall unit with a reconfigurable device. The firewall unit needs a large amount of register for the timing adjustment of packets. Using the registers is the cause of power consumption. In this paper, to solve the problem of power consumption, the firewall unit has developed by using wave-pipelining technique and detailed delay time for the technique is analyzed.

**Keywords: Firewall, Mobile devices, Wave-pipelining, Delay time analysis.**

## 1. INTRODUCTION

A firewall is indispensable to prevent unauthorized access to computers connected to computer networks. It is needed in not only server computers but also mobile devices such as a PDA and a smartphone. The mobile devices have a feature that operates with a small battery. Therefore, the firewall can be used without sacrificing this feature is required. As a firewall that achieves these demands, logic-based firewall circuits on FPGA have developed [1].

Additionally, the authors have developed H-HIPS (Hardware- and Host-based Intrusion Prevention System) [2]. It has been implemented on FPGA (Field programmable gate array), and its detection units have been achieved by logic circuits that can be reconfiguration.

H-HIDS has an IDL (Intrusion Detection Logic). The IDL of NIC that plays in part the function of network-base IDS should process the packet analysis more than three layers of OSI (Open Systems Interconnection) layer model. After manufacturing custom design VLSI processor cannot change hardware, it cannot add a limitation to a specific protocol and a new function. Therefore, it cannot be used for IDL. Then, the hardware processing that is only software becomes possible by using FPGA as for IDS. The TCP/IP Flow Monitor circuits were achieved with FPGA for such reasons [3].

The firewall contributes to not only the prevention of an illegal access but also reduction of unauthorized detection circuits. Moreover, the firewall enables the power control to aim at the power reduction of the detection circuits of H-HIPS and networks and mobile communications systems. .

The novelty of the circuits is to be able to achieve wave-pipelined operations without changing the composition of the circuits. Because the wave-pipelined circuits execute the pipeline operations without using registers, the operations lead a low-power operation and a high-speed operation. However, the firewall circuits are only the circuits for the judgment of an illegal port and

don't have a shift register necessary to operate as the firewall.

Conventional shift register necessary to operate as the firewall is not mounted. Building conventional shift register into it influences the operation speed of the firewall circuits. In addition, to use a lot of registers, those use has the problem to increase the power consumption. Therefore, wave-pipelined shift register has been developed. Building into the firewall unit and the delay time analysis of the whole circuits is needed.

This paper is organized as follows. Section 2 presents wave-pipelining. Then, Section 3 describes reconfigurable firewall units by using wave-pipelining. The results of delay time analysis presented on Section 4. In Section 5, the conclusions are made.

## 2. WAVE-PIPELINING

Not only high clock frequency but also low power dissipation can be obtained at the same time by wave-pipelining [4]-[6]. It exploits high throughput combinational logic blocks in which as many as data are launched unless they conflict. Although wave-pipelining was attempted to the entire region of a processor, it was viewed pessimistically because it requires removing general registers as well as pipeline registers from processors. It seems hard to eliminate general registers playing fundamental roles in sequential circuits. The insufficient power of CAD tools so far developed is another reason why wave-pipelines have been applied out of processors.

Accordingly, design and evaluation methods for wave-pipelines have not yet established well compared with those for conventional pipelines. Mostly wave-pipelines have been so far applied to simple unifunctional circuits such as adders, multi-pliers, counters, and DRAM. Regarding multifunctional wave-pipelines, a wave-pipelined ALU has recently appeared [7]-[10].

Then, a microprocessor developed by using wave-pipelines in part appeared [10]. It is 14-segment ULTRASPARC-III whose second and third instruction fetch segments have been wave-pipelined. Another example is an asynchronous wave-pipeline, though it is not compatible with conventional processors. Wave-pipelined CRC that is sequential circuits has been developed by us [11].

The signal path of combinational circuit is uneven of delay time. The most high-speed signal in one group in a clock has the possibility to collide with the slow signal in just before clock. The problem was solved to wave-pipelined combinational circuit shown in Figurer 1 (a) by the delay time of all signal paths is brought close at the delay time of critical path.

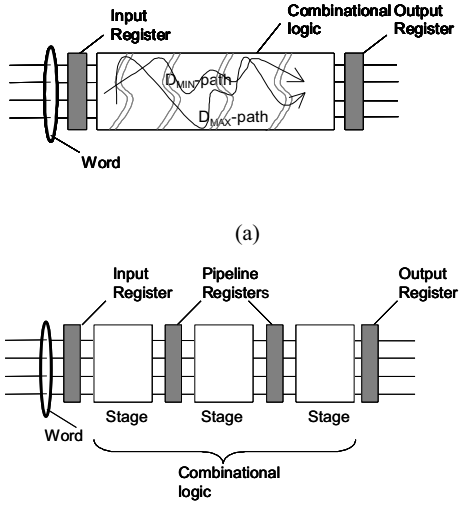
The relation between the clock cycle and delays is obtained as follows [6].

$$T_{CK} > ( D_{MAX} - D_{MIN} ) + T_{OV}. \quad (1)$$

Here,

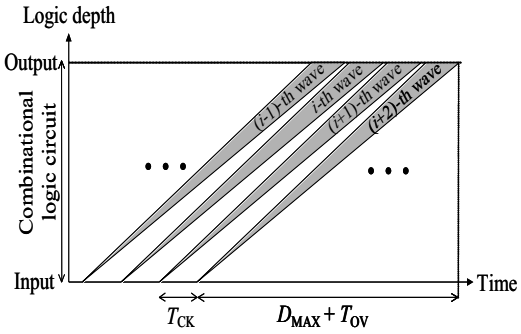
$T_{CK}$: Clock cycle time

$T_{OV}$: Overhead time




Figurer 1. Synchronization of pipelines. (a) Wave-pipeline. (b) Conventional pipeline.

From Eq. (1), $D_{MAX} - D_{MIN}$ should be close to 0 as much as possible in order to obtain minimum $T_{ck}$. One of solution to satisfy this requirement can be conceived from Figure 2 that shows relation between time and logical depth.




Figurer 2. Wave model of Figurer 1 (a).

## 3. RECONFIGUREURABLE FILEWALL

Figure. 3 shows the outline of the firewall for H-HIPS. In the last work of [], the controlled ports are for using a mobile computing, and they are at least needed. Table I is the controlled ports. Because the firewall unit is developed by FPGA, the change of ports is very easy.

TABLE I

CONTROLLED PORTS

| Function | Port Number | Binary |
|---|---|---|
| NOP | 0 | 0000000000000000 |
| SMTP | 25 | 0000000000011001 |
| DNS | 53 | 0000000000110101 |
| HTTP | 80 | 0000000001010000 |
| POP3 | 110 | 0000000001101110 |

TABLE II

DEVELOPMENT ENVIRONMENT

| | |
|---|---|
| CPU | Intel Core 2 Duo E6600 (2.4GHz) |
| Memory | 2G Bytes |
| OS | Micorosoft Windows XP Pro. SP3 |
| Logic synthesis | Altera Quartus II V8.0 |
| Simulator | Mentor Graphics ModelSim Altera 6.1g |
| FPGA Device | Altera Cyclone EP1C20F400C7 |

Figure 4 shows synthesized circuits by using the development environment of Table II. Maximum delay time of the circuits is 17.9 ns. The circuits can operate at 50 MHz by conventional operations. And, Minimum delay time is 12.3 ns. According to Eq. (1), they can operate at 100MHz by wave-pipelined operations. The gate-level simulations are executed for confirming wave-pipelined operations. Figure 5 shows conventional operations and Figure. 6 shows wave-pipelined operations. According to the results of Figure 6, 100 MHz operations are confirmed.

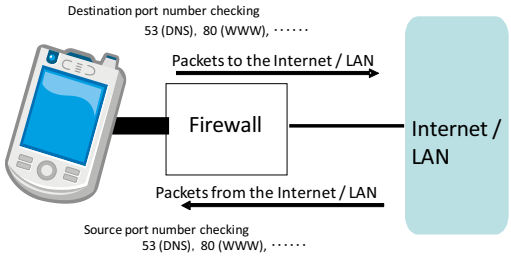



Figure 3. Firewall for H-HIPS.

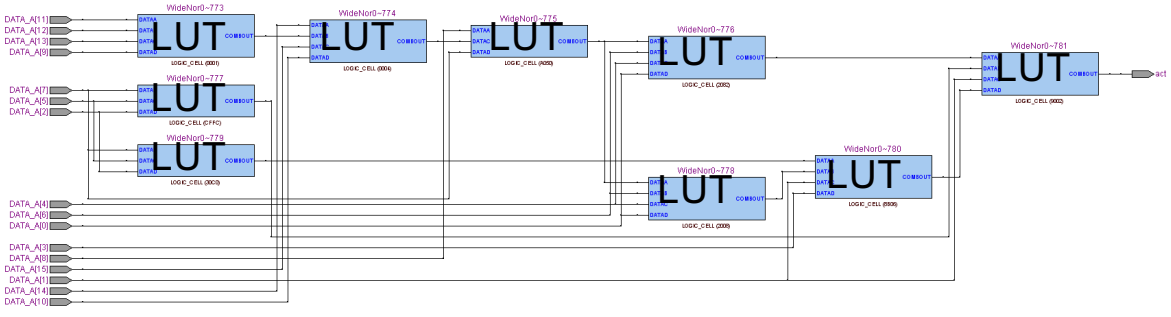



Figure 4 Firewall Unit.

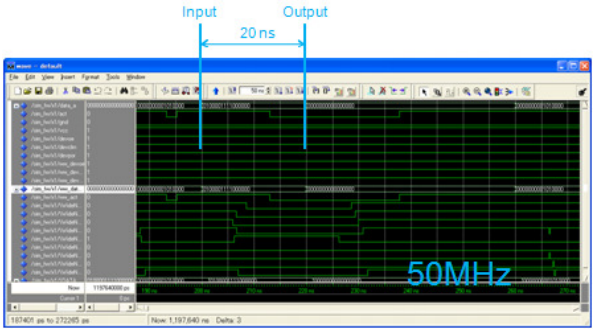



Figure 5. Conventional Operations (50MHz).

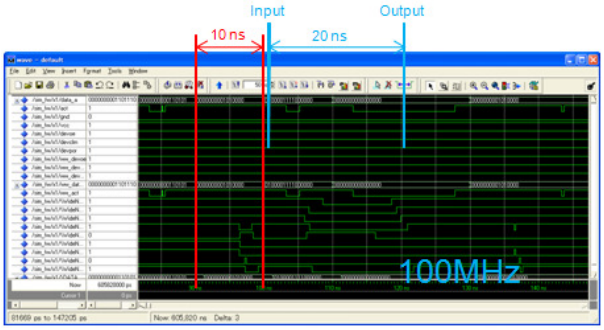



Figure 6. Wave-Pipelined Operations (100MHz).

*A. Shift Register for the Firewall Circuits*

The firewall circuits are only the circuits for the judgment of an illegal port number. They are necessary for each observation of a source port number and a destination port number. Therefore, the authors develop the firewall unit shown in Figurer 7.

Detection results in the source port number and the destination port number obtained by the firewall circuits are input to an OR gate. Then, the value of the OR gate is input to a flag register. When an illegal port number is detected by the flag register, the packet including the illegal port number must be destroyed. Accordingly, a shift register is needed for the packet.

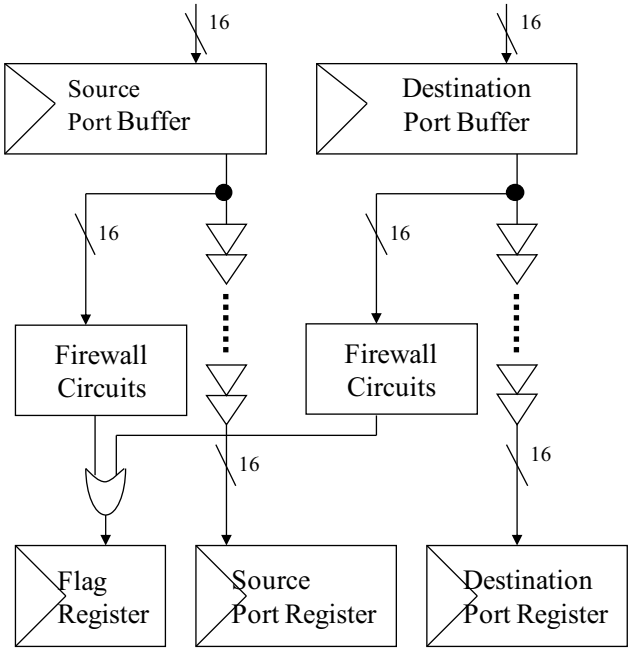



Figure 7. Firewall Unit.

However, the firewall circuits operate by wave-pipelined operations. That is, because operation speed of a conventional shift register is slower than that of the firewall circuits, the conventional shift register cannot be used. Therefore, a wave-pipelined shift register that is composed by buffers is developed and is built in.

*B. Operation Speed*

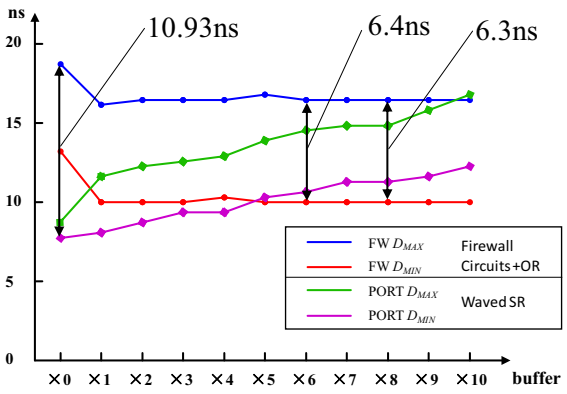
The delay time of the firewall unit shown in Figure 7 is analyzed. The development environment of Table II is used for the analysis. As a result, the maximum delay time is 8.80 ns and the minimum delay time is 3.33 ns. In the development of the firewall unit, buffer insertion and logic synthesis are executed. As a result, the different maximum delay time from the results of Sec. III is obtained. From Eq. (1), the firewall unit operates at 180 MHz.

## 4. DELAY TIME ANALYSYS

In this section, the optimized number of the buffers for the delay time adjustment is found by delay time analysis. Figure 8 shows relation between the number of buffers and delay time. It was made from the following.

- Dividing the circuit of Figure 7 into Firewall Circuits with OR gate and Waved Shift Registers.
- Buffer numbers of Waved Shift Registers are adjusted.
- The maximum delay time and the minimum delay time of each circuits are obtained by using the CAD.

According to Eq. (1), it is dominant in the point at the speed that the difference at the maximum delay and the minimum delay time is minimum. When the buffer number is 8 in Figure 8, it is shown that it is minimum in 6.3ns. On the other hand, delay time of 6 buffers is 6.4ns. In this case, 6 buffers are selected from the point of the area.




Figurer 8. Relation between the Number of Buffers and Delay Time.

## 5. CONCLUDING REMARKS

In this paper, the authors analyzed the optimal number of buffers for wave-pipelining operations of Reconfigurable Firewall Unit. In the case of 8 buffers, the results of the analysis shows maximum operating speed. However, 6 buffers were selected from the point of the area. Future works are detailed evaluations by using a packet data.

## ACKNOWLEDGMENT

This work has been supported in part by Grant-in-Aid for Young Scientists (B) (23700068) from Japan Society for the Promotion of Science (JSPS), Japan.

## REFERENCES

[1] Tomoaki Sato, Syuya Imaruoka, and Masa-aki Fukase, “ReconFigureurable Firewall Unit by Wave-Pipelined Operations,” proc. of ISPACS 2008, pp. 449-452, 2009.

[2] Tomoaki Sato, Kazuhira Kikuchi, Syuya Imaruoka, and Masa-aki Fukase, “DoS Attack Analysis for H-HIPS,” Proc. of IMETI, Vol. II, pp. 110-115, 2008.

[3] David V. Schuehler and John W. Lockwood, “TCP Splitter: A TCP/IP Flow Monitor in ReconFigureurable Hardware,” IEEE Micro, Vol. 23, No. 1, pp. 54-59, 2003.

[4] L. Cotton, “Maximum rate pipelining systems,” Procs. AFIPS Spring Joint Computer Conference, pp. 581-586, 1969.

[5] F. Klass and M. J. Flynn, “COMPARATIVE STUDIES OF PIPELINED CIRCUITS,” Stanford University Technical Report, No. CSL-TR-93-579, July 1993.

[6] W. P. Burleson, M. Ciesielski, F. Klass, and W. Liu, “Wave-Pipelining: A Tutorial and Research Survey,” IEEE Trans. on Very Large Scale Integration (VLSI) Systems, Vol. 6, No. 3, pp. 464-474, Sept. 1998.

[7] T. Sato, M. Fukase, and T. Nakamura, “Performance analysis of a wave-pipelined ALU,” Technical Report of IEICE, CPSY 2000, Vol. 100, No. 20, pp. 1-6, 2000.

[8] M. Fukase, T. Sato, R. Egawa, and T. Nakamura, “Scaling up of Wave Pipelines,” THE FOURTEENTH INTERNATIONAL CONFERENCE ON VLSI DESIGN, Jun. 2000.

[9] M. Fukase, T. Sato, R. Egawa, and T. Nakamura, “Breakthrough of Superscalar Processors by Multifunctional Wave-Pipelines,” Proc. of 9th NASA Symposium on VLSI Design, pp. 6.3.1-6.3.17, Nov. 2000.

[10] M. Fukase, T. Sato, R. Egawa, and T. Nakamura, “Designing a Wave-Pipelined Vector Processor,” Proc. of the Tenth Workshop on Synthesis and System Integration of Mixed Technologies, pp. 351-356, Oct. 2001.

[11] Tomoaki Sato, Phichet Moungnoul, keisuke Saito, Masa-aki Fukase, "Wave-Pipelined CRC Circuits for Wireless

Broadband Systems Based on W-CDMA," Proc. of ICESIT2010, pp. 100.1-100.4, 2010.

[12] Tomoaki Sato, Kei Ito, Keisuke Saito, Phichet Moungnoul and Masa-aki Fukase, "Development of a shift register for Firewall Circuits by Wave-Pipelined Operations," Proc. of 2010 International Workshop on Information Communication Technology, pp. w4c-1-1-w4c-1-4, 2010.

# A new imaging technique based on the polarization of light

**Isabella Chiara BUSCEMI**
**LiSSi Laboratory, Université Paris Est**
**Créteil, France**

**and**

**Steve GUYOT**
**LiSSi Laboratory, Université Paris Est**
**Créteil, France**

## ABSTRACT

This work is the first presentation of a new imaging technique which enables near real time multispectral acquisition of the so called "Degree Of Polarization" (DOP) in polarimetry. In this context DOP is the normalized difference between two orthogonal states of polarization. Essentially it is used as the contrast element to determine the difference of light propagation in turbid media for different polarization. This approach is well known considering linear and/or circular polarizations, but the new experimental setup also permits the acquisition of DOP for all the possible elliptic states of polarization allowed by using of nematic liquid crystals.

KEYWORDS: Mueller's matrix, polarimetric imaging, Degree of Polarization (DOP)

Correspondence: *isabella-chiara.buscemi@univ-paris-est.fr*

## 1. INTRODUCTION

The aim of the study of the optical properties of a medium is to define some quantitative besides qualitative information and, among the four light characterization factors, the polarization has been chosen to highlight contrast as just proposed by Jacques [5], Guyot [2, 6] and Anastaziadou [7].

In this sense the calculation of the DOP for each input polarization state is done in order to appraise the comprehension of the mechanism of scattering. In particular it is considered in all his types (such as Rayleigh, Gans, Mie) [8]. We consider the variation of the input polarization state such as linear, circular (the most studied) and elliptic to obtain a complete dynamic characterization of the scattering media response.

## 2. METHOD

Considering the scattering media, the classical approach in polarimetry consists in the acquisition of a complete Mueller's matrix (MM)[4]. Several polarimetric properties of the media like diattenuation, retardance or depolarization can be studied by MM [1, 3, 4]. In the special case of biological media which is our interest, the main problem of the Mueller's formalism is the acquisition time (typically from a few seconds to few tens of seconds). Problems and consequences are largely discussed in [2].
The required experimental setup is shown in Figure 1.

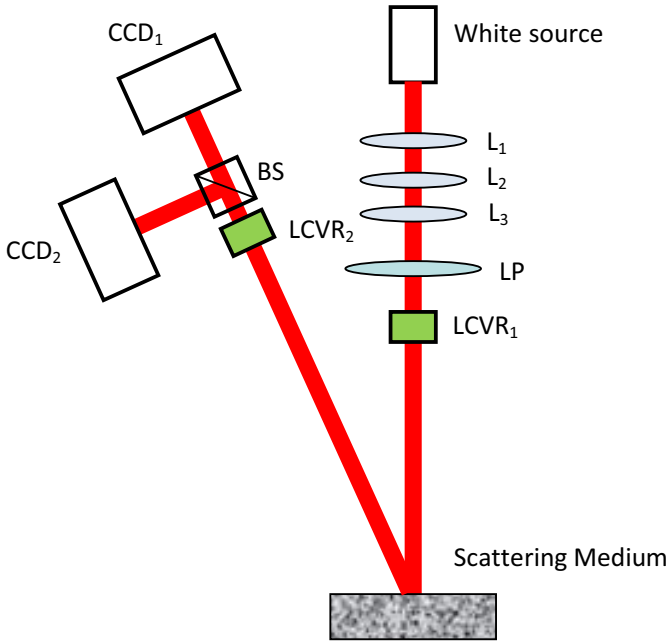



Fig. 1 shows a schematic layout of the system.

The illumination part of the set-up comprises a halogen lamp (Olympus CLH-SC 150W), a fiber bundle with its output at the focus of an aspherical condenser $L_1$ (LA-1027 Thorlabs $f_1$ = 35 mm) followed by an achromatic lens $L_2$ (LB-1471 Thorlabs, $f_2$ = 50 mm) whose focal point coincides with the condenser's one. It should be noted that this system provides a homogeneous illumination over the field of view. A second achromatic lens $L_3$ ($f_3$ = 100 mm) images the source on the sample. The Polarization State Generator (PSG) was developed with a linear polarizer LP (LPVIS-100MP Thorlabs) followed by a nematic liquid crystals variable retarders $LCVR_1$ (LCR-1VIS Thorlabs) which is remotely controlled by a power supply. Particularly, it allows the continuous variation of the beam polarization. By this method, it is possible to generate all the feasible polarizations from linear state to elliptic and circular ones.

Moreover, as demonstrated in Figure 1, we keep an angle of 15° between the optical PSG axis and the PSA (Polarizer State Analyzer) axis. This angle allows us to prevent specular reflection on the medium and so imaging artifacts [5].

The PSA is composed of a second $LCVR_2$ lens (coupled in charge and orthogonal fast axis with the first one) and a polarized beam splitter cube BS which allows to separate the two perpendicular polarization components ($I_{//}$ and $I_{\perp}$) of the output intensity on the two CCD cameras (Pixelfly QE, 1392X1024 pixels, Cooke). We give an example of both the results in Figure 2, while in Figure 3 the obtained calculated DOP image.

It is important to note that a calibration of the complete system is required for the perfect alignment of the several components on one hand, but on the other hand also for the validation of the multispectral acquisition. Indeed, for this second point, it is necessary to determine the retards introduced by the LCVR vs voltage for the different input wavelength. In particular this is connected with the multispectral behavior of the used nematic liquid crystal that has to be especially characterized.

Moreover, considering the biological media and the well known “therapeutic windows”, it seems important to be able to better know the differences of interaction considering the wavelength of the incident beam, its polarization and scattering conditions.

Note that all the system is piloted by a routine in Matlab with SDK function.

The scientific case for this setup is to study light propagation in turbid medium, as illustrated in our previous works [2, 3, 6, 7], but the complete exploitation of light in imagery remains an important theoretical problem.

Fig. 2: images with perpendicular components of polarization.

Fig. 3: DOP image obtained by images in Figure 2.

## 3. CONCLUSION

To summarize, in this paper we report the first validation of our new imaging techniques. This experimental setup allows the near real time multispectral acquisition of the so called "DOP" in polarimetry. This "Degree Of Polarization" can be obtained with controlled ellipticity (from linear through elliptic to circular DOP) of the incident beam.

We demonstrate all the accuracy of our setup during this conference and present our first results on human skin and more precisely concerning application for the diagnosis in dermatology.

## 4. REFERENCES

[1] S. Y. Lu and R. A. Chipman, **Interpretation of Mueller matrices based on polar decomposition**, J. Opt. Soc. Am. A 13, 1106-1113 (1996)

[2] S. Guyot, M. Anastasiadou, E. Delechelle, A. De Martino, **Registration scheme suitable to Mueller matrix imaging for biomedical applications**, Optics express, Vol. 15, Issue 12, 7393-7400, (2007)

[3] R. Ossikovski, A. De Martino, S. Guyot, **Forward and reverse product decomposition of depolarizing Mueller matrices**, Opt. Lett. Vol. 32, No. 6, pp. 689-691 (2007)

[4] R. M. A. Azzam, **A simple Fourier photopolarimeter with rotating polarizer and analyzer for measuring Jones and Mueller matrices**, Opt. Comm. Vol. 52, No. 2, pp. 137-140 (1978)

[5] S. L. Jacques, J. C. Ramella-Roman, K. Lee, **Imaging skin pathology with polarized light**, J. Biomed. Opt. 7(3), pp. 329-340 (2002)

[6] S. Guyot, C. Vézien, B. Clairac and B. Fontas, **An exploitation of the Mueller's formalism: differentiation of irradiate collagene from non-irradiate one after polarimetric measurements**, Opt. Appl. Vol. 34, No. 2 (2004)

[7] M. Anastasiadou, S. Ben Hatit, R. Ossikovski, S. Guyot and A. De Martino, **Experimental validation of the reverse polar decomposition of depolarizing Mueller matrices**, J. E. Opt. Soc., Vol. 2, 07018 (2007)

[8] A. A. Kokhanovsky, **Polarization Optics of random media**, Springer (2003)

# Diagnosis of Esophageal Cancer via Correlation Coefficient of Optical Density

**Shao-Ting CHEN, I-Jen HSU***
**Department of Physics, Chung Yuan Christian University, Chungli 32023, Taiwan**

**and**

**Chiang-Ni LU, Max Ti-Kuang HOU**
**Department of Mechanical Engineering, National United University, Miaoli 36003, Taiwan**

**and**

**Jang-Ming LEE**
**Department of Surgery, National Taiwan University Hospital, Taipei 10002, Taiwan**

## ABSTRACT

The currently used techniques for the detection of esophageal cancer usually rely on pathological examination after sectioning and ultrasound imaging. Optical methods have the advantages of better spatial resolution and sensitivity with minimal invasion. Therefore, the development of optical techniques will be helpful for the diagnosis of esophageal cancer at the early stage. In this research, we propose and develop an optical system for the detection of optical densities of biological samples in the visible range. Both the optical densities of normal and cancerous esophageal tissues from three independent patients as well as that of cultured esophageal cells were measured. After normalization of the spectra and calculation of the correlation coefficient of each pair of spectra with developed algorithms, the distribution of correlation coefficients was mapped for the identification of normal and cancerous tissues. We have also developed an algorithm which can quantitatively evaluate the possibility of a site of tissue to be cancerous. The results show significant distinguishability between normal and cancerous tissues and consistency with pathological examinations. Since the utilized light is in the visible range that is easy to be delivered in optical fibers, this technique is suitable for *in vivo* detection in the integration with endoscopes.

**Keywords:** Esophageal Cancer, Optical Diagnosis, Spectrum Analysis, Optical Density, Correlation Coefficient.

## 1. INTRODUCTION

According to the report announced by the Department of Health in Taiwan, cancer has been the leading one among the top ten causes of death in Taiwan in the past two decades. The types of cancers led to death in recent years include lung cancer, liver cancer, colorectal cancer, female breast cancer, gastric cancer, oral cancer, prostate cancer, cervical cancer, esophageal cancer and pancreatic cancer, where the ranking of esophageal cancer gradually raises from the twelfth to the ninth from 1996 to 2008. Meanwhile, the mortality rate in patients with esophageal cancer rises from 3.6% to 6.2%. In addition, the median age of death of the top two types of cancers, the lung cancer and liver cancer, were delayed from 69 to 73 and from 64 to 68, respectively. However, the median age of death of esophageal cancer changed from 66 to 58—a phenomenon of youth oriented. Therefore, it is imperative to develop the techniques for the diagnosis and treatment of esophageal cancer at the early stage.

There are many reasons lead to esophageal cancer. In Taiwan, betel nut chewing is one of the main risks of esophageal cancer [1]. The tobacco and alcohol habits will result in a substantial increase of chances of esophageal cancer [2]. In addition, consumption of hot drinks or soup resulting in allergy and damage to the esophagus, head or other digestive organs had been suffering from cancer and other genetic factors are also leading causes of esophageal cancer. Furthermore, patients with Barrett's esophagus are at high risk of suffering from esophageal cancer [3]. In addition to the dietary habits and environmental factors which are relative to the occurrence of esophageal cancer, the researchers indicated that the incidence of esophageal cancer and genetic background of individuals is also relevant. For example, congenital variation (polymorphism) of genes such as p53 or XRCC1 can also affect individual risk of suffering from esophageal cancer [4, 5]. On the other hand, these differences in genetic background may also affect the results of treatment for patients. For example, the angiotensin converting enzyme gene can affect the patient in the risk of the occurrence of pulmonary complications after surgery of esophageal cancer [6]. The mutation existed in the normal population of glutamic acid conversion enzyme gene will have a meaningful impact to the survival after treatment of the patients [7]. Exposure to the environmental toxicants will cause gene damage and affect the stability of p53 [8, 9].

The currently used techniques for diagnosis of esophageal cancer include computer tomography (CT) [10], magnetic resonance imaging (MRI) [11], chromoendoscopy [12], acoustic microscopy [13] and narrow-band imaging [14]. As the spectroscopic techniques have the advantages of non-invasion and low-destruction, it is widely studied and used in biomedicine. For example, infrared spectroscopic technique can be used for distinguishing between the various pharmaceutical drugs, analysis of drug purity and crystal structure, and test of

the effects of different drugs [15]. Infrared spectroscopy has also been used in clinical diagnosis such as the diagnosis of traumatic intracranial hematoma in order to reduce the renewal of the injury [16] and detection of cervical precancerous lesions with the combination of Raman spectroscopy [17]. In addition to infrared spectroscopy, fluorescence spectroscopy is commonly used in biomedical technology, such as the use of fluorescence spectroscopy for analysis of protein interactions [18] and the use of fluorescent labeling for tumor tracking and study of cell differentiation.

There are various techniques of the diagnosis and treatment of esophageal cancer. However, the effect of endoscopic detection for esophageal lesions at the early stage is not clear. Therefore, many techniques were developed to improve the ability of endoscopic detection for early esophageal lesions including spectroscopy. Spectroscopy provides a rapid and noninvasive diagnosis by detecting the signal through the interaction between light beam and tissues [19]. The relative stuies include the use of laser excited autofluorescence to diagnose different stages of esophageal lesions [20], the use of measurements of back-scattered spectrum to diagnose Barrett's esophagus [21], assessment of the capability of endoscopic measurement of Raman spectrum [22] and detection of Barrett's esophagus and esophageal adenocarcinoma with endoscopic optical coherence tomography [23].

In this research, we developed an optical system for the detection of optical densities of biological samples in the visible range. Both the optical densities of normal and cancerous esophageal tissues from three independent patients as well as that of cultured esophageal cells were measured. After normalization of the optical densities and calculation of the correlation coefficient of each pair of optical densities with developed algorithms, the distribution of correlation coefficients was mapped for the identification of normal and cancerous tissues. We have also developed an algorithm which can quantitatively evaluate the possibility of a site of tissue to be cancerous. The results show significant distinguishability between normal and cancerous tissues and consistency with pathological examinations. Since the utilized light is in the visible range that is easy to be delivered in optical fibers, this technique is suitable for *in vivo* detection in the integration with endoscopes.

The purpose of this research is the measurement of the transmissive absorption spectra of the normal and cancerous human esophageal tissues as well as cultured esophageal cancer cells in the visible range with a self-designed spectrum measurement system. We also developed an algorithm for analysis and process of measured spectral information and can be used for quantitative pathology diagnosis. The results show significant distinguishability between normal and cancerous tissues and consistency with pathological examinations. Since the utilized light is in the visible range that is easy to be delivered in optical fibers, this technique is suitable for *in vivo* detection in the integration with endoscopes.

## 2. MATERIALS AND METHODS

In this research, the optical densities of human esophageal specimens were used as the basis for evaluation of the possibility of a tissue to be cancerous. Three groups of normal and cancerous esophageal specimens from independent patients as well as cultured human esophageal cancer cells were under investigation. In order to quantitatively analyze the possibility of an esophageal specimen to be cancerous, we developed an algorithm for statistical processing of measurement results. The optical density is defined as

$$A(\lambda) = \log \frac{I_0(\lambda)}{I(\lambda)}, \tag{1}$$

where $I_0(\lambda)$ is the spectrum of incident light, $I(\lambda)$ is the spectrum of transmitted light through the specimen. In our research, measurements of the optical densities at different positions of each esophageal specimen were performed. The correlation coefficients between the optical densities were discussed to distinguish the optical densities of normal and cancerous tissues. The distribution of the correlation coefficients was then obtained.

After the measurements of the optical densities, the correlation coefficient between each pair of normalized optical densities was calculated by use of two kinds of definitions of the correlation coefficients:

$$c_{ij}^{(1)} = 1 - \frac{\int |A_i(\lambda) - A_j(\lambda)| d\lambda}{\int [A_i(\lambda) + A_j(\lambda)] d\lambda} \tag{2}$$

and

$$c_{ij}^{(2)} = \int \mathrm{Min}[A_i(\lambda), A_j(\lambda)] d\lambda. \tag{3}$$

In addition to the study of the differences of the optical densities between the normal and cancerous esophageal tissues, we also calculated the distribution of the correlation coefficients to identify the standard optical densities corresponding to normal and cancerous esophageal tissues which can then be used to quantitatively determine the possibility of a tissue to be cancerous.

After obtaining the correlation coefficient of each pair of optical densities, the standard deviation of the distribution of the correlation coefficients relative to each optical density was calculated. The two optical densities corresponding to the largest two standard deviations in the distributions of relevant correlation coefficients were identified as the standard optical densities of normal and cancerous tissues, respectively. The possibility of a tissue to be cancerous was then calculated with the following formula:

$$P = \frac{\frac{c_{iC}}{c_{NC}} - \frac{c_{iN}}{c_{NN}}}{\frac{c_{CC}}{c_{NC}} - \frac{c_{iN}}{c_{NN}}}, \tag{4}$$

where $c_{ij}$ denotes the correlation coefficient defined in Eq. (2) and Eq. (3). The subscript $N$ and $C$ correspond to the standard optical densities of normal and cancerous tissues, respectively.

## 3. RESULTS

The optical densities of the three groups of specimens as well as the cultured human esophageal cells were measured with a homemade spectrum measurement system where a broadband xenon lamp (SUTTER, Lambda LS) is used as the light source. The transmitted signal is delivered with an optical fiber and detected with a spectrometer (Ocean Optics, HR4000) after a neutral-density filter.

In the experiments, 19 optical densities were measured at different sites of the three groups of normal and cancerous esophageal specimens. Optical densities of three groups of cultured esophageal cancer cells were also measured. The labels of 1 to 22 corresponding to different specimens and groups of cells are listed in Table. 1. After the calculation with the procedure mentioned in the previous section, the standard optical densities of normal (label 4) and cancerous (label 6) tissues were identified, respectively. The possibility of each optical density to be cancerous one was calculated by use of Eq. (4). Table 2 shows the results of calculated possibility of each optical density to be cancerous one.

Table 1 Labels of optical densities corresponding to the esophageal tissues and cells.

| Label of patient | Tissue type | Label of optical density |
|---|---|---|
| 1 | normal | 1-4 |
| | cancerous | 5-8 |
| 2 | normal | 9-11 |
| | cancerous | 12-14 |
| 3 | normal | 15-17 |
| | cancerous | 18, 19 |
| Cultured cells | cancerous | 20-22 |

Table 2 Possibility of optical density to be cancerous one.

| Label of patient | Tissue type | Label of optical density | Possibility of cancer |
|---|---|---|---|
| 1 | normal | 1 | 1.3% |
| | | 2 | 0.9% |
| | | 3 | 0.7% |
| | | 4 | 0.0% |
| 1 | cancerous | 5 | 57.4% |
| | | 6 | 100% |
| | | 7 | 56.4% |
| | | 8 | 95.5% |
| 2 | normal | 9 | 7.5% |
| | | 10 | 32.9% |
| | | 11 | 6.2% |
| 2 | cancerous | 12 | 3.9% |
| | | 13 | 35.5% |
| | | 14 | 3.8% |
| 3 | normal | 15 | 3.2% |
| | | 16 | 5.3% |
| | | 17 | 3.4% |
| 3 | cancerous | 18 | 11.8% |
| | | 19 | 46.7% |
| Cultured cells | cancerous | 20 | 31.6% |
| | | 21 | 45.3% |
| | | 22 | 40.3% |

Due to Table 2, the average value of the possibility of cancer of the 10 optical densities of normal tissues is 6.14%, where the average value of the possibility of cancer of the 9 optical densities of cancerous tissues is 45.67%. The average value of the possibility of cancer of the 3 optical densities of cultured esophageal cancer cells is 39.07%. The results show significant distinguishability between normal and cancerous tissues and consistency with pathological examinations.

## 4. CONCLUSIONS

The normalized optical densities show significantly distinguishable features between normal and cancerous tissues. There are strong correlation between optical densities of cancerous tissues as well as that between cancerous tissue and cultured cancer cells. Similar results were found in normal esophageal tissues. In the results of quantitative evaluation of the possibility of cancer for a specific site of the tissue, we can see that most of the results indicate the possibility for that tissue to be cancerous are very small when the tissue was pathological examined to be normal. On the other hand, the results indicate that there are high possibilities to be cancerous at some site for the tissues pathological examined to be cancerous. Similar results were found in the optical densities of cultured esophageal cells. The only exception is the optical density labeled as 10 which was suspected to be cancerous in the calculation. However, the calculated value is still not very large and more data is needed for further confirmation. Some optical densities in the cancerous tissues refer to small possibility to be cancerous. It is reasonable to expect that not the whole area in the tissue is cancerous. Therefore, the results in our experiments are basically consistent with pathological examinations. With the quantitative evaluation of the distribution, the contour of cancerous region can be identified under a two dimensional scanning. The use of visible light results in the possibility for this technique to be integrated in endoscope for *in vivo* applications.

## 5. ACKNOWLEDGMENT

This research was supported by the National Science Council of Republic of China under the grant no. NSC 99-2627-B-033-003.

## 6. REFERENCES

[1] Y. C. Ko, Y. L. Huatig, C. H. Lee, M. J. Chen, L. M. Litt, and C. C. Tsai, “Betel quid chewing, cigarette smoking and alcohol consumption related to oral cancer in Taiwan,” **Journal of Oral Pathology and Medicine**, Vol. 24, 1995, pp. 450-453.

[2] I. C. Wu, C. Y. Lu, F. C. Kuo, S. M. Tsai, K. W. Lee, W. R. Kuo, Y. J. Cheng, E. L. Kao, M. S. Yang, and Y. C. Ko, “Interaction between cigarette, alcohol and betel nut use on esophageal cancer risk in Taiwan,” **European Journal of Clinical Investigation**, Vol. 36, 2006, pp. 236-241.

[3] A. J. Cameron, C. T. Lomboy, M. Pera, and H. A. Carpenter, “Adenocarcinoma of the esophagogastric junction and Barrett's esophagus,” **Gastroenterology**, Vol. 109, 1995, pp. 1541-1546.

[4] J. M. Lee, Y. C. Lee, S. Y. Yang, W. L. Shi, C. J. Lee, S. P. Luh, C. J. Chen, C. Y. Hsieh, and M. T. Wu, “Genetic polymorphisms of p53 and GSTP1, but not NAT2, are

associated with susceptibility to squamous-cell carcinoma of the esophagus," **International Journal of Cancer**, Vol. 89, 2000, pp. 458-464.

[5] J. M. Lee, Y. C. Lee, S. Y. Yang, P. W. Yang, S. P. Luh, C. J. Lee, C. J. Chen, and M. T. Wu, "Genetic polymorphisms of XRCC1 and risk of the esophageal cancer," **International Journal of Cancer**, Vol. 95, 2001, pp. 240-246.

[6] J. M. Lee, A. C. Lo, S. Y. Yang, H. S. Tsau, R. J. Chen, and Y. C. Lee, "Association of angiotensin-converting enzyme insertion/deletion polymorphism with serum level and development of pulmonary complications following esophagectomy," **Annals of Surgery**, 2005, Vol. 241, pp. 659-665.

[7] J. M. Lee, M. T. Wu, Y. C. Lee, S. Y. Yang, J. S. Chen, H. H. Hsu, P. M. Huang, S. W. Kuo, C. J. Lee, and C. J. Chen, "Association of *GSTP1* polymorphism and survival for esophageal cancer," **Clinical Cancer Research**, Vol. 11, 2005, pp. 4749-4753.

[8] J. M. Lee, T. Y. Liu, D. C. Wu, H. C. Tang, J. Leh, M. T. Wu, H. H. Hsu, P. M. Huang, J. S. Chen, C. J. Lee, and Y. C. Lee, "Safrole-DNA adducts in tissues from esophageal cancer patients: clues to areca-related esophageal carcinogenesis," **Mutation Research**, Vol. 565, 2005, p. 121.

[9] J. M. Lee, C. T. Shun, M. T. Wu, Y. Y. Chen, S. Y. Yang, H. I. Hung, J. S. Chen, H. H. Hsu, P. M. Huang, S. W. Kuo, and Y. C. Lee, "The associations of p53 overexpression with p53 codon 72 genetic polymorphism in esophageal cancer," **Mutation Research**, Vol. 594, 2006, pp. 181-188.

[10] H. M. Nob, E. K. Fishman, D. F. Bliss, and P. S. Calhoun, "CT of the esophagus: spectrum of disease with emphasis on esophageal carcinoma," **Radio Graphics**, Vol. 15, 1995, pp. 1113-1134.

[11] B. Kumbasar, "Carcinoma of esophagus: radiologic diagnosis and staging," **European Journal of Radiology**, Vol. 42, 2002, pp. 170-180.

[12] M. I. F. Canto, "Esophageal cancer screening by chromoendoscopy," **Techniques in Gastrointestinal Endoscopy**, Vol. 1, No. 2, 1999, pp. 50-52.

[13] K. S. Wickramasinghe, P. T. Chandrasoma, and P. A. N. Chandraratna, "Detection of Barrett's epithelium by acoustic microscopy," **Ultrasound in Medicine and Biology**, Vol. 28, No. 2, 2002, pp. 203-207.

[14] H. Mizuno, K. Gono, S. Takehana, T. Nonami, and K. Nakamura, "Narrow band imaging technique," **Techniques in Gastrointestinal Endoscopy**, Vol. 5, No 2, 2003, pp. 78-81.

[15] G. N. Kalinkova, "Infrared spectroscopy in pharmacy," **Vibrational Spectroscopy**, Vol. 19, No 2, 1999, pp. 307-320.

[16] C. S. Robertson, S. P. Gopinath, and B. Chance, "Use of near-infrared spectroscopy to identify traumatic intracranial hematomas," **Journal of Biomedical Optics**, Vol. 2, 1997, pp. 31-41.

[17] U. Utzinger, D. L. Heintzelman, A. Mahadevan-Jansen, A. Malpica, M. Follen, and R. Richards-Kortum, "Near-Infrared Raman spectroscopy for in vivo detection of cervical precancers," **Applied Spectroscopy**, Vol. 55, No 8, 2001, pp. 955-959.

[18] Y. Yan and G. Marriott, "Analysis of protein interactions using fluorescence technologies," **Current Opinion in Chemical Biology**, Vol. 7, 2003, pp. 635-640.

[19] W. K. Song and B. C. Wilson, "Optical detection of high-grade dysplasia in Barrett's esophagus," **Techniques in Gastrointestinal Endoscopy**, Vol. 7, No 2, 2005, pp. 78-88.

[20] M. Panjehpour, B. F. Overholt, T. Vo-Dinh, R. C. Haggitt, D. H. Edwards, and F. P. Buckley, "Endoscopic fluorescence detection of high-grade dysplasia in Barrett's esophagus," **Gastroenterology**, Vol. 111, No 1, 1996, pp. 93-101.

[21] M. B. Wallace, L. T. Perelman, V. Backman, J. M. Crawford, M. Fitzmaurice, M. Seiler, K. Badizadegan, S. J. Shields, I. Itzkan, R. R. Dasari, J. Van Dam, and M. S. Feld, "Endoscopic detection of dysplasia in patients with Barrett's esophagus using light-scattering spectroscopy," **Gastroenterology**, Vol. 119, 2000, pp. 677-682.

[22] M. G. Shim, L. M. Song, N. E. Marcon, and B. C. Wilson, "In vivo near-infrared Raman spectroscopy: demonstration of feasibility during clinical gastrointestinal endoscopy," **Photochem Photobiol**, Vol. 72, No 1, 2000, pp. 146-150.

[23] S. Jackle, N. Gladkova, F. Feldchtein, A. Terentieva, B. Brand, G. Gelikonov, V. Gelikonov, A. Sergeev, A. Fritscher-Ravens, J. Freund, U. Seitz, S. Schroder, and N. Soehendra, "In vivo endoscopic optical coherence tomography of esophagitis, Barrett's esophagus, and adenocarcinoma of the esophagus," **Endoscopy**, Vol. 32, No 10, 2000, pp. 750-755.

# Modeling Frequency Response of Photoacoustic Cells using FEM for Determination of N-heptane Contamination in Air: Experimental Validation

**Lars DUGGEN, Mihaela ALBU, Morten WILLATZEN, and Horst-Günter RUBAHN**
**NanoSYD, Mads Clausen Institute, University of Southern Denmark**
**Alsion 2, DK-6400 Sønderborg, Denmark**

## ABSTRACT

We briefly present the basic principle of the photoacoustic effect in gases. We present the equations and boundary conditions governing the acoustic field generated by the absorption of a modulated laser beam. We solve these equations using Finite Element Methods and compare the results with experiment. We find that apparently there are effects not taken into account in the classic acoustic theory as we find a larger damping than theory predicts. However, we see that these effects have significant negative influence on the quality factor of the cell and thereby the performance limit.

**Keywords:** Photoacoustic effect, acoustic damping, FEM, n-heptane contamination.

## 1. INTRODUCTION

The photoacoustic effect was discovered by Alexander Graham Bell in 1880 and has drawn considerable scientific and engineering attention after the 1960's in connection with the invention and availability of lasers [1]. Its broad range of applications include photoacoustic spectroscopy, with the advantage of applicability to solids (also soft tissue) as well as fluids. This diversity makes it feasible for non-invasive medical applications, biological, and environmental applications [2, 3, 4].

Here we focus on the measurement of n-heptane concentrations in air, which is a model system for an application towards measurement of oil contaminations in air. For experimental convenience we chose to conduct experiments on n-heptane as it shares many critical properties (as optical absorption spectrum) with oil without having the drawback of sticking to the surfaces. Thus, compared to using oil, the photoacoustic cell is easy to clean after experiments.

Since the generated acoustic signal might be rather small, it is convenient to make use of acoustic resonance as a natural amplifier of the signal. In order to design the cell according to desired requirements such as resonance frequency (such that it fits to the microphone and laser driver specifications), mathematical modeling is a crucial tool. Mathematical models for the photoacoustic effect have been developed since the middle of the 1970's, started by a theoretical description of the photoacoustic effect with solids [5, 6, 7]. Regarding gas-photoacoustics there exist several works considering the theory behind [8], specifically loss mechanisms [7, 9, 1]. However, many of these works are limited to rather simple geometries, being a limiting factor for the effective design of a photoacoustic cell. There have also been works concerning the simulation of photoacoustic cells using Finite Element Methods [10, 11, 12], however there still are problems modeling the loss mechanisms correctly.

Despite using a rather simple geometry, in principle covered by the existing theory, we will present a slightly more general model and solve it using the

Finite Element Method (FEM), comparing it with the experiment in order to find a reliable method to model more complex photoacoustic cells using FEM. Furthermore we will discuss experimental results and analysis of these.

## 2. THEORY

The basic theory behind the photoacoustic effect can briefly be described as follows, focusing on the photoacoustic effect in gas. The contaminated air (i.e. synthetic air and n-heptane) is lead into a closed cell as depicted in Figure 1. Then a laser beam is sent through the cell. The wavelength of the laser is tuned such that it fits to a contaminant absorption peak. When being absorbed, molecular rotational/vibrational states are excited and these states are subsequently relaxed by molecular collisions, essentially heating up the gas in the small volume around the particular molecule that absorbed laser light. Finally the laser intensity is modulated (in this case sinusoidally) and the heating will cause thermal and pressure waves throughout the gas mixture. The pressure wave is then measured by a microphone. The amplitude of the acoustic signal depends linearly on the strength of the laser and the concentration of the contaminant (up to possible saturation limits at very low concentrations).

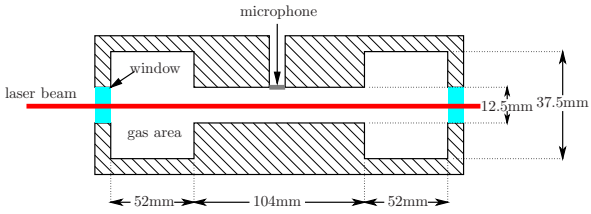



Figure 1: Cross section of the cylindrical photoacoustic cell used. The cell is made of stainless steel, gold coated while the windows are made of BK7 glass.

The photoacoustic effect can be described theoretically by fundamental thermodynamic laws regarding the absorption of the laser light as heat energy input [13, 14]. Manipulations as well as assumptions about a small viscosity and small heat transfer coefficient (such that only first order terms need to be accounted for) together with small temperature and pressure deviations, yields the Helmholtz equation in acoustic pressure $p$ reading

$$\nabla^2 p + k^2 p = -i\omega \frac{P_0 \gamma \epsilon}{T_0 C_p c^2}. \tag{1}$$

Here it is assumed that the heat input function $\epsilon$ is a monofrequency input such that general time dependence can be obtained by a Fourier transform of the time dependent input. Here, $p$ denotes the acoustic pressure (i.e. pressure change generated by the laser), $C_p$ is the heat capacity at constant pressure, $\gamma$ denotes the ratio of specific heats, $c$ is the speed of sound, and $P_0, T_0$ are the equilibrium pressure and temperature, respectively. The value of $k$ is the wavenumber $\omega/c$, where we pick the lossless wavenumber as the classical bulk loss effects are negligible [11].

The boundary condition in radial coordinates reads [11]

$$\left.\frac{\partial p}{\partial r}\right| r_R = i\omega\rho_0 \left[\frac{1}{z} - \frac{i+1}{\rho_0 c}\left((\gamma-1)\frac{J_1\left(\sqrt{\frac{i\omega}{l_h c}}R\right)}{J_0\left(\sqrt{\frac{i\omega}{l_h c}}R\right)}\sqrt{\frac{l_h\omega}{2c}} + \frac{k_t^2 c^2}{\omega^2}\frac{J_1\left(\sqrt{\frac{i\omega}{l_v c}}R\right)}{J_0\left(\sqrt{\frac{i\omega}{l_v c}}R\right)}\sqrt{\frac{l_v\omega}{2c}}\right)\right] p, \tag{2}$$

where $z$, $\rho_0$, $J_n$, $k_t$ denote acoustic impedance of the wall material, equilibrium mass density, $n$-th order Bessel function, and transverse wavenumber (see discussion in [11, 13]), respectively. Furthermore $l_h, l_v$ denote thermal and viscous loss terms given by

$$l_h = \frac{K}{\rho_0 C_p c}, \tag{3}$$

$$l_v = \frac{\mu}{\rho_0 c}, \tag{4}$$

where $K$ is the heat conductivity and $\mu$ is the bulk viscosity.

This equation is solved for the given geometry using Comsol Multiphysics 3.5® as two-dimensional problem with the Cartesian differential operators transformed into their polar counterparts.

We solve both the case where all losses are taken exactly as above (sim. w/o fit) and the case where we fit the loss parameters in this simulation to experimental results (sim. with fit). Finally the amplitude of the simulation with fit is scaled such that it corresponds to the experimental amplitude. The ratio between the two simulated results is kept constant in order to compare the two simulations. This indicates the real quality of the cell compared to the theoretical (i.e. ideal) quality of the cell of given dimensions.

## 3. EXPERIMENTAL SETUP

The cell used is depicted in Figure 1. The photoacoustic signal is detected by a fibre optic microphone "Optimic 4110 FA" from Optoacoustics. This ultra low noise microphone is designed for monitoring very weak sound signals in selected bandwidths and has a sensitivity of 5 V/Pa. It has a resonance frequency of 1.65 kHz and a $-3$ dB conformity from 1.40 to 1.80 kHz. The electrical signal is furthermore filtered and amplified with a "Kemo BenchMaster" 8.41 filter/amplifier unit with an adjustable band pass filter (8 pole Butterworth filter $-48$ dB/octave) from Kemo Limited. The refurbished signal is connected to a lock-in amplifier (model SR 830 from Stanford Research Systems) to extract the correlated amount from the signal. A calibrated commercial instrument, "1314 Photoacoustic Multi-gas Monitor" from Innova Air Tech Instrumennts, was used as a reference to monitor the contamination level of n-heptane. Each acoustic frequency is excited and measured for three seconds.

The contaminated gas samples are prepared by initially introducing a small amount of n-heptane (typically around 20 ppm) into synthetic air and measuring the concentration by the reference device. Subsequently, after each measurement series covering the acoustic frequencies of interest, this sample is diluted to half of the previous concentration, thus giving a reliable concentration.

The measurements are then fitted to the following function, based on the response of the damped oscillator, i.e. on the same fundamental principle resulting in damped harmonic vibrations:

$$p_{fit} = \frac{A}{\sqrt{\omega_s^4 m^2 + \omega_s^2 (R_m^2 - 2sm) + s^2}}, \qquad (5)$$

where $\omega_s = 2\pi(f - f_s)$ with $f$ being the acoustic frequency in Hz and $f_s$ is a fitting parameter. Furthermore $s, m$, and $R_m$ are fitting parameters roughly corresponding to bulk modulus, mass density, and allover damping, respectively. However, these parameters are only fitted once as they represent cell and gas properties which do not change significantly (assuming temperature and pressure are kept almost constant). Finally, for each measurement set the parameter $A$ is fitted, being the indicator of signal strength.

This fitting function has shown to work reasonably well and is quite robust to single signal outliers. Also we have tried several polynomial fitting functions of degrees up to seven. The harmonic fitting function showed to have the least accumulated error $\sum_{n=1}^{n_{meas}} (p_{fit,n} - p_{meas,n})^2$, where $p_{fit,n}$ are the respective pressures obtained by the fitting function at a certain frequency (denoted by index $n$, i.e. $f = f_0 + n \cdot \Delta f$) and $p_{meas}$ are the corresponding measured pressures, while $n_{meas}$ simply denotes the amount of measured pressures.

## 4. RESULTS

We have used two different cell designs; one a described above (figure 1) and one as described in [12].

The result for a single concentration measurement obtained by the prototype described in [11] is shown in Figure 2. We have found that, using the simulation with fitted loss parameters, we obtain very good agreement with experiments, indicating the feasibility of this model.

Results for the acoustic signal at a concentration of 14 parts-per-million (ppm), using the cell design described here, are shown in Figure 3.

We find, as before [11], that the damping considered in the boundary condition given above is not sufficient to describe the actual acoustic signal. Note that for the dotted curve we have used standard values for the thermodynamic properties of air only; we

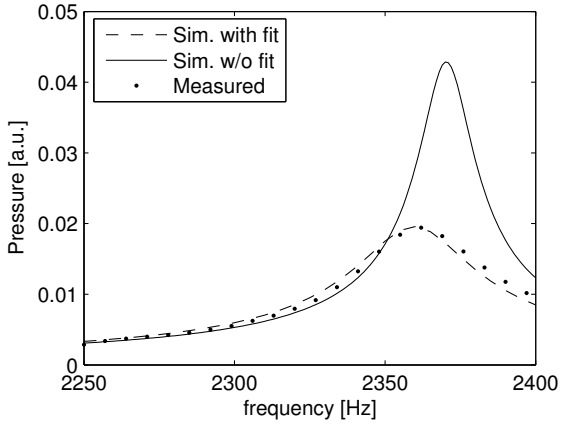



Figure 2: Measured and simulated pressure for a photoacoustic cell of total length 143 mm, a resonator diameter of 15 mm and a buffer diameter of 35 mm [12]

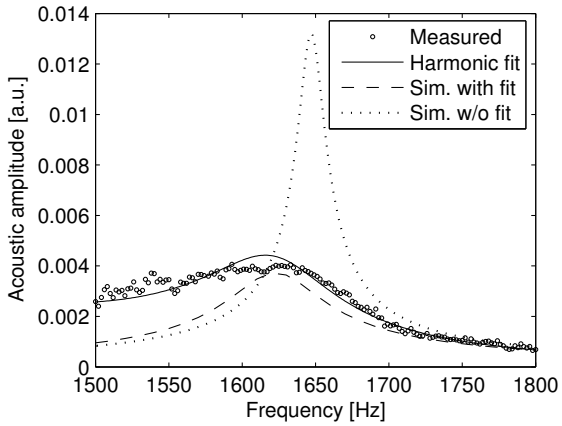



Figure 3: Acoustic signal with n-heptane concentration of 14 ppm at different frequencies

have only scaled the amplitude of the curve as the measured signal is also subject to electronic amplification. However, we find that an artificial increase of the boundary effects, obtained by scaling of the $l_h$ and $l_v$ parameters, leads to a better agreement. This is seen as the dashed curve in Figure 3, where we conduct the same simulation with the imaginary terms in equation (2) multiplied by 3.4. Nevertheless, there seem to be some effects in the cell that our model does not capture, as the shape of the simulated response does not fit the measured response very well.

The amplitude of the acoustic signal is shown in Figure 4, where the amplitude is seen to depend on the n-heptane concentration somewhat linearly, although with some deviations. This indicates the importance of a complete understanding of the acoustic effects apparent in the cell in order to find the origin of these deviations and remove them as far as possible in the cell design.

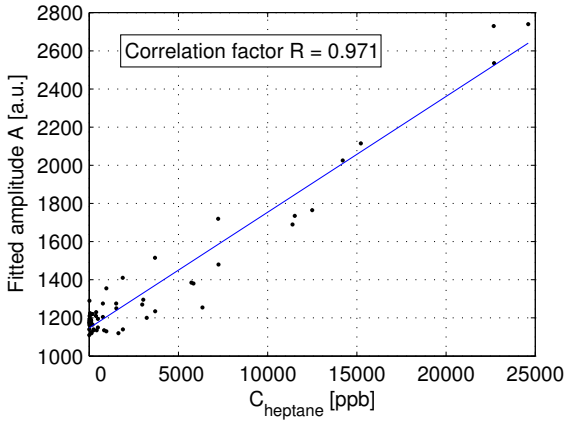



Figure 4: Measured pressure amplitudes at resonance for different n-heptane concentrations, using the cell depicted in Figure 1

## 5. CONCLUSIONS

We have briefly presented the basics of the photoacoustic effect and have employed the theory to simulate the acoustic behavior of the cell using Finite

Element Methods.

We have conducted experiments using synthetic air contaminated with n-heptane and found that a modified version of a damped oscillator resonance is a quite reasonable fitting function, being more stable than high-order polynomials and giving the least accumulated error.

We have found a reasonably good linear behavior of signal strength vs. n-heptane concentration, although there are deviations.

Because of the fact that we obtain quite good agreement with one prototype (Figure 2) and the reasonable agreement only on one edge with the other prototype (Figure 3), we believe it is important to examine the frequency response further, measuring other resonance frequencies as well as changing other parameters and reconsidering details of the model.

## 6. ACKNOWLEDGMENTS

We gratefully acknowledge the financial support from the Danish National Advanced Technology Foundation.

## 7. REFERENCES

### References

[1] A. Miklos, P. Hess, and Z. Boszoki. *Rev. Sci. Instrum.*, 72:1937, 2001.

[2] C. Haish and R. Niessner. *Spectroscopy Europe*, 14:10, 2002.

[3] F.H.C. Bijnen, J. Reuss, and F.J.M. Harren. *Rev. Sci. Instrum.*, 67:2914, 1996.

[4] B.T. Cox, S. Kara, S.R. Arridge, and P.C. Beard. *J. Acoust. Soc. Am.*, 121:3453, 2007.

[5] A. Rosencwaig and A. Gersho. *Journal of Applied Physics*, 47:64, 1976.

[6] F.A. McDonald and G.C. Wetsel. *Journal of Applied Physics*, 49:2313, 1978.

[7] R.D. Kamm. *Journal of Applied Physics*, 47:3550, 1976.

[8] G.J. Diebold and P.J. Westerveld. *J. Acoust. Soc. Am.*, 84:2245, 1988.

[9] J.B. Mehl and M.R. Moldover. *J. Chem. Phys.*, 74:4062, 1981.

[10] B. Baumann, B. Kost, H. Groninga, and M. Wolff. *Rev. Sci. Instrum.*, 77:044901, 2006.

[11] L. Duggen, N. Lopes, M. Willatzen, and H.-G. Rubahn. *Intern. Journ. of Thermophys.*, page 1, 2010 (online first).

[12] L. Duggen, R. Frese, and M. Willatzen. *Jour. of Phys.: Conf. Series*, 214:012036, 2010.

[13] P.M. Morse and K.U.Ingard. *Theoretical Acoustics*. Princeton Univ. Press, Princeton, NJ, 1986.

[14] L.E. Kinsler, A.R. Frey, A.B. Coppens, and J.V. Sanders. *Fundamentals of Acoustics*. Wiley & Sons Inc., Hoboken, NJ, 1999.

# Supersolution focused spot using a radially polarized beam and a filter with continuous amplitude and discrete phase modulation

**Kun Huang and Yongping Li**

**Department of Optics and Optical Engineering, University of Science and Technology of China, Hefei, Anhui, 230026,China**

**Abstract**

A filter with continuous amplitude and discrete phase modulation is introduced. For a focused radially polarized using the filter and a tight focusing lens, a supersolution focused spot and a strong longitudinally polarized field in the focal region are obtained in theory. The tight focusing system is approximated by Wolf and Richards' theory. The filter introduced here enriches the family of generating an optical "needle" of longitudinally polarized field.

**Keywords:** Polarization; Longitudinal field; Tight focusing.

## 1. Introduction

Recently, the radially polarized beam has been gaining great attention for its novel properties. The special polarization makes the radially polarized beam have wide applications in many fields, i.e. optical trapping, microscopy, optical date store, particle acceleration, second-harmonic generation, molecule orientation imaging and quantum communication. Especially, the ability of obtaining the supersolution spot using the radially polarized beam and a high numerical-aperture (NA) lens has been investigated in the past few years. Dorn et al. demonstrated experimentally that a $0.16\lambda^2$ (its FWHM is $0.45\lambda$) spot was achieved by focusing a radially polarized beam using a 0.9NA lens with an annular aperture [1]. Wang et al. introduced the multi-belt binary-phase elements into the system of focusing the radially polarized beam and attained an optical "needle" with the size of $0.43\lambda$ (FWHM) and long depth of focus [2]. Huang et al. utilized the focusing properties of the radially polarized beam to improve the method of designing the multi-belt phase elements [3]. Zhan et al. suggested a filter with discrete amplitude and phase modulation for the purpose of obtaining an optical "needle" with the supersolution size [4]. Here, we provide a filter with continuous amplitude and discrete phase modulation to achieve the supersolution spot in the focal plane when focusing a radially polarized beam with a high NA lens. The focused spot with the size of $0.40\lambda$ is achieved in the free space. To the best of our knowledge, the $0.40\lambda$ spot in the free space is the smallest spot that has been reported using the traditional optical elements so far. When we change the structure of the filter, an optical "needle" of longitudinally polarized field is obtained in the focal region. The filter designed in this paper enriches the family of generating the optical "needle" with longitudinally polarized field.

## 2. Theory

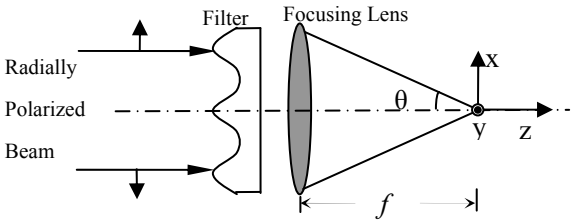



Fig. 1. Schematic of focusing of a radially polarized beam with a filter and a high NA lens. The filter can modulate the amplitude and phase of light. $\theta$ denotes the angle between the convergent ray and the optical axis. $f$ is the focus of the lens.

For a tight focusing system, the fields in the focal region can be approximated by the Richards and Wolf's theory. In this paper, the NA of the focusing lens is 0.90 and the media in the focal region is air. For a focused radially polarized beam, the fields in the focal region

contain the radial ($E_r$) and longitudinal ($E_z$) component fields.

$$E_r(\rho,\varphi,z) = A\int_0^{\theta_{\max}} P(\theta)\cos^{1/2}\theta\sin(2\theta) \cdot J_1(k\rho\sin\theta)e^{ikz\cos\theta}d\theta, \quad (1)$$

$$E_z(\rho,\varphi,z) = 2iA\int_0^{\theta_{\max}} P(\theta)\cos^{1/2}\theta\sin^2\theta \cdot J_0(k\rho\sin\theta)e^{ikz\cos\theta}d\theta, \quad (2)$$

where $\rho,\varphi,,z$ are the polar coordinates, $A$ is a constant, $J_n$ is the n-order Bessel function of the first kind, $P(\theta)$ is the complex amplitude of light that is incident on the focusing lens. In general, $P(\theta)=E_0(\theta)\bullet T(\theta)$, where $E_0(\theta)$ is the amplitude of the radially polarized beam and $T(\theta)$ is the transmission function of the filter. For simplicity, we assume that the intensity of the radially polarized beam is uniform in this paper, that is, $E_0(\theta)=1$. From equations (1) and (2), one can see that, the longitudinal component field has non-zero on-axis intensity while the radial component field has null on-axis intensity. The longitudinal component field dominates the focal region. Therefore, the spot size in the focal plane is determined by the size of the longitudinal component field. In addition, the outmost of the incident radially polarized beam contributes most to the longitudinal component field in the focal region. Therefore, considering the focusing properties, one can easily conclude that much attention should be paid to modulate the outmost radially polarized beam. We propose a filter whose transmission function is given by:

$$T(r) = M\sin(2\pi Fr)\Big/\sqrt{(\alpha R/r)^2-1}, \quad (3)$$

where $R$ is the radius of the entrance pupil of the lens, $r$ is the radial position (polar coordinate) of the beam ($r\le R$), $F$ is the modulation frequency, $\alpha$ is a constant (slightly larger than 1, $\alpha$=1.01 in this paper) and $M^{-1} = \left|\sin(2\pi Fr)\Big/\sqrt{(\alpha R/r)^2-1}\right|_{\max}$. For a lens that obeys the sine condition, $\sin\theta=r/f$. The transmission function of the filter contains two factors: $1\Big/\sqrt{(\alpha R/r)^2-1}$ and $\sin(2\pi Fr)$. The factor

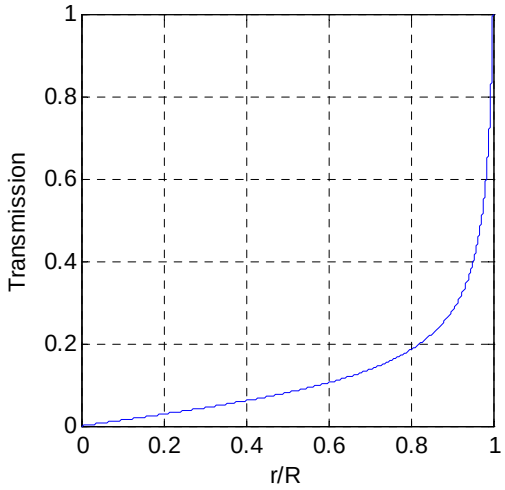



Fig. 2. Transmission function of the "high-pass" filter.

$1\Big/\sqrt{(\alpha R/r)^2-1}$ works as a "high-pass" filter, whose transmission function is displayed in Fig. 2. The "high-pass" filter maintains the outmost beam and depresses the inmost beam. In addition, the factor $\sin(2\pi Fr)$ provides the sine amplitude and discrete phase modulation. The discrete phase is associated with the sign of the sine factor $\sin(2\pi Fr)$. The modulation frequency $F$ can be an arbitrary constant or a function of position $r$.

## 3. Results and Discussion

We achieve the numerical solutions of equations (1) and (2) by programming in MATLAB 7.0 (Mathworks). Considering the properties of the "high-pass" filter, one can easily expect an optical "needle" with longitudinally polarized filed in the focal region. In order to obtain the "needle", one can adjust the structure of the filter by the method of changing the modulation frequency $F$. When searching the filter, we use the method in Ref. 3 to evaluate the beam quality of the optical "needle". As mentioned above, the modulation frequency $F$ can be variable.

### 1) Frequency *F* is an arbitrary constant

The factor $\sin(2\pi Fr)$ is the sine function of the position $r$. The variable $F$ just changes the filter's frequency. After optimizing the frequency $F$, one can find the right filter for generating an optical "needle" of longitudinally

polarized field in the focal region. Here, we suggest a frequency $F$=0.71. The transmission function of the filter with frequency $F$=0.71 is shown in Fig.3 (a). It is necessary to note that the negative value of transmission function stands for the phase delay "$\pi$" of the filter. Therefore, the filter introduced in this paper has the modulation of continuous amplitude and discrete phase. Fig. 3(b, c and d) in the x-z plane give the fields in the focal region. One can clearly see that a strong optical "needle" of longitudinally polarized field is achieved by combining the tight focusing lens and the suggested filter with $F$=0.71. Fig. 3(b) displays a weak radial field component and a strong longitudinal field component is shown in Fig.3 (c). Comparing Fig. 3(c) and (d), we can conclude that the longitudinally polarized field dominates in the focal region. The depth of focus of the optical "needle" is about 3.5λ and the lateral size of the "needle" is 0.478λ (FWHM). Compared with the "needles" in Ref. 2 and 3, the "needle" designed in this paper has larger lateral size. However, the structure of the filter in this paper is equivalent to a three-belt binary phase element in Ref. 2 and 3. Considering the number of belts (5 belts in Ref. 2 and 4 belts in Ref.3), we have advantage over the number of belts of binary phase element that is used to create an optical "needle" of longitudinally polarized field. In addition, the "needle" designed in this paper have no sidelobes (strong axial intensity [4]).Therefore, from the point of the "needle" quality, the "needles" in Ref. 4 and this paper may have wider applications than those in Ref. 2 and 3.

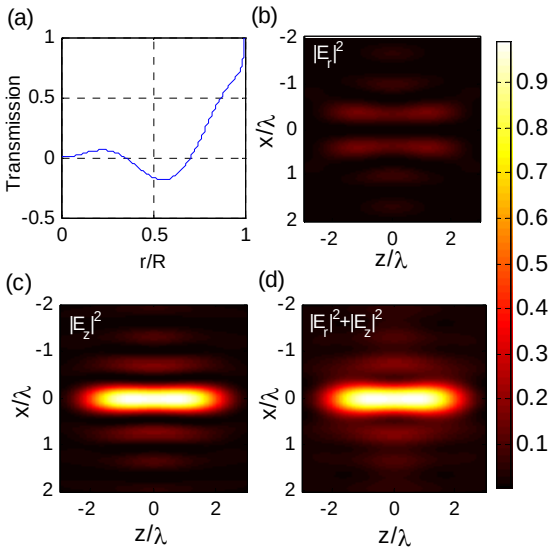



Fig.3. The transmission function of the filter with frequency $F$=0.71 and the corresponding fields ($E_r$ and $E_z$) in the focal region in the x-z plane. (a) Transmission function of the filter. (b) Radial component field. (c) Longitudinal component field. (d)Total field. The normalized intensity is shown in (b), (c), and (d).

**2) Frequency *F* is a function of position *r***

Although frequency $F$ can be an arbitrary function of $r$, we choose the frequency $F(r)$=C • arcsin($r/f$) for the purpose of introducing the stronger modulation of the outmost light. The filter can be adjusted by changing the variable C. However, the filter is different from that with a given frequency. Fig. 4(a) gives the transmission of a filter with variable C=0.78. Although the filter in Fig.4 (a) also introduces the modulation of phase, the modulation frequency becomes larger as the radial position $r$ increases. The filter in Fig.3 (a) has a given modulation frequency. When C=0.78, a spot with subwavelength size can be obtained in the focal plane. The FWHM of the spot is 0.40λ. The intensity profile of the spot is displayed in Fig. 4 (b).

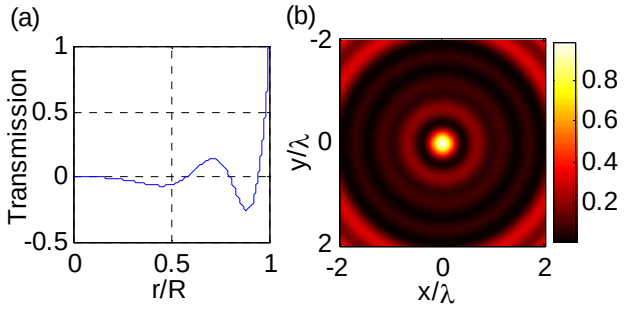



Fig. 4 (a) Transmission function of a filter with variable C=0.78. (b) The intensity profile of the corresponding spot in the focal plane. The normalized intensity is shown here.

When C=4.48, the filter gives the strong modulation and the transmission is shown in Fig. 5(a). After the modulation of the filter, a strong longitudinally polarized "needle" in the focal region is generated. Fig. 5 (b) and (c) demonstrate the radial and longitudinal component field in the focal region, respectively. The total field (Fig. 5(d)) in the focal region displays the profile of the longitudinally polarized "needle". The lateral size of the "needle" is 0.43λ.

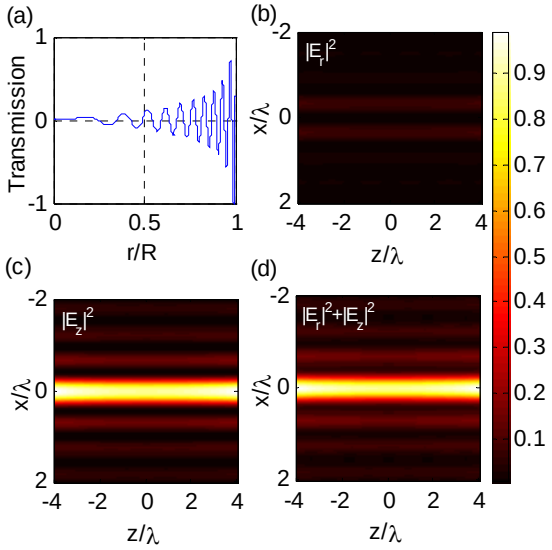



Fig. 5 The transmission function of the filter with frequency $C$=4.48 and the corresponding fields ($E_r$ and $E_z$) in the focal region in the x-z plane. (a) Transmission function of the filter. (b) Radial component field. (c) Longitudinal component field. (d)Total field. The normalized intensity is shown in (b), (c), and (d).

In addition, because we constrain the filter to have a "high-pass" profile, the fields in the focal region are dominated by the longitudinally polarized field and an optical "needle" can be easily realized. In fact, if one designs a filter with "band-pass" profile, the radially polarized field in the focal region will increase. When the radial and longitudinal component fields are equal, an interesting optical "bubble" can be generated in the focal region. Of course, many filters can be used to modulate the radially polarized beam and many new results can also be obtained.

## 4. Conclusion

We introduced a filter with continuous amplitude and discrete phase modulation. In theory, using the filter and a tight focusing lens to focus a radially polarized light with uniform intensity, we obtained several longitudinally polarized "needles". The filter suggested in this paper enriches the family of creating a longitudinally polarized "needle". Although the filter has the amplitude and phase modulation, it can be easily realized by the spatial light modulation (SLM) with amplitude and phase modulation experimentally.

## References

[1] R. Dorn, S. Quabis, and G. Leuchs, "Sharper focus for a radially polarized light beam". Physical Review Letters 91, 233901 (2003).

[2] H. Wang, L. Shi, B. Lukyanchuk, C. Sheppard, and C.T. Chong, "Creation of a needle of longitudinal polarized light in vacuum using binary optics," Nature Photonics 2, 501–505 (2008).

[3] K. Huang, P. Shi, X. L. Kang, X. B. Zhang, and Y. P. Li, Design of DOE for generating a needle of a strong longitudinally polarized field. Optics Letters 35, 965-967 (2010).

[4] Jiming Wang, Weibin Chen, and Qiwen Zhan. "Engineering of high purity ultra-long optical needle field through reversing the electric dipole array radiation". Optics Express 18 (21): 21965-21972

# A visual prosthesis for long term epiretinal stimulation in blind patients with retinitis pigmentosa

**Matthias Keserue**
**Department of Ophthalmology,**
**University Medical Center Hamburg-Eppendorf**
**Hamburg, Germany**

**and**

**Gisbert Richard**
**Department of Ophthalmology,**
**University Medical Center Hamburg-Eppendorf**
**Hamburg, Germany**

## Abstract

**Purpose:** To evaluate the performance of the first wireless Retinal Implant System in blind patients with retinitis pigmentosa. Objective of the study is to evaluate safety and efficacy as well as stability and optimal stimulus parameters for single electrode as well as simple pattern stimulation. Further objective is to control the wireless signal transmission for both data and energy and to test the feasibility of the surgical implantation method and the toleration of the implant by the eye over time. **Methods:** In a prospective study, 3 subjects with a visual acuity of hand movement or less received an epiretinal implant with 49 stimulation electrodes (IMI Intelligent Medical Implants GmbH). On predefined postoperative dates, stimulation sessions of 3-4 hours were performed, including threshold measurements, amplitude modulations, motion and pattern perception tests. All results were verified by repeating testing. ICG, fluorescein angiograms (HRA, Heidelberg Engineering) and OCT investigations (Stratus OCT and Cirrus OCT, Carl Zeiss Meditec) were also achieved. **Results:** A stable implant positioning of the electrode array over the whole followup period was identified. Fluorescein angiograms showed no signs of vascular leackage or tissue proliferation. OCT data proved a close proximity between the retina and the electrode film. No retinal tissue damage was observed. The provoked visual perceptions were pleasant. Reported perceptions were predominantly circles, yellow or white in colour. Threshold measurements showed safe and stable current levels. Point-to-point-discrimination was successful in all subjects. Motion as well as pattern perceptions showed success rates up to 100%. The patients were able to describe structures similar to letters. Data and energy transmission worked well. **Conclusions:** The implant allows safe and stable stimulation parameters in combination with good clinical results. This seems promising for upcoming implantations. The completely wireless data and energy transmission into the implant allows totally unrestricted eye movement and is therefore suitable as a long-term solution.

**Keywords:** epiretinal implant, retinitis pigmentosa, retinal prosthesis, retinal stimulation

## 1. Introduction

Retinitis pigmentosa (RP) and macular degeneration especially age related macular degeneration are the most common diseases, caused by a continuous loss of photoreceptor cells. With an incidence of 1 in 3500 live births and approximately 1.5 million people affected worldwide, RP is the leading cause of inherited blindness. RP is evoked by a genetic disorder and is characterized by a progressive visual field loss and decline of visual acuity leading to complete blindness in most patients. However the inner part of the retina, the bipolar and ganglion cells, remains intact in RP, so that the higher order neurons are still capable to process information to the brain and visual cortex. Thus the replacement of the degenerated photoreceptors by an electric stimulator might be a therapeutic option in RP and other degenerative retinal diseases. This idea is not new. The first reports on eliciting phosphenes by electric stimulation of the visual system date back to the 18th century [1]. Since then many different working groups have tried to elicit visual perceptions by eletrical stimulation. This development was especially driven by the success of cochlear implants. Two main approaches emerged and proved successful in retinal stimulation systems; the epiretinal and the subretinal approach [2]. From a surgical point of view, the epiretinal approach is easier to perform and is therefore associated with a lower incidence of peri- and postoperative complications.
We have been developing an epiretinal stimulator, that can be surgically implanted using standard vitrectomy techniques and is wirelessly recieving stimulation data and energy [3,4]. This paper reports on the clinical and stimulation results of a chronic epiretinal implantation of this wireless epiretinal prosthesis in blind patients with RP.

## 2. Materials and Methods

*The Epiretinal implant*

The device used in the study is developed and manufactured by IMI Intelligent Medical Implants GmbH, Bonn, Germany. It consists of an implanted component, the retinal stimulator, and several external components. The implantable retinal stimulator

(Figure 1) consists of a carrier onto which various micro-electronic components are mounted. The electronic components are used to receive data and energy as well as to stimulate the retina via micro-electrodes. The carrier is a flexible circuit film with built-in conducting paths. The embedded high frequency receiver coil of this carrier provides the required energy for the stimulation electronics. The coil is embedded in a supporting structure and positioned in the temporal area outside the eyeball in close proximity to an external sending coil which transmits the necessary energy. The infrared-receiver receives data containing visual information from an external transmitter via an infrared optical link. The infrared-receiver encodes the optical signals into electric stimulation currents and sends them to the retinal stimulator-chip. This chip processes the data and generates stimulation patterns for the electrode array placed on the retina. The array is composed of 49 platinum stimulation electrodes and is positioned on the epiretinal side of the retina in the area of the macula. The electrodes are in direct contact with the retina and activate the retinal nerve cells via electrical stimulation currents. The Retina stimulator is coated with Parylene C. The material was deposited on the device by plasma activation with following vapour deposition. The material is resistant against acids, blood, moisture and hydrocarbons. A LED indicator located near the array serves to control the proper functioning of the device by visual inspection within the eye. The encapsulated outer part of the retinal stimulator is sutured onto the sclera. The intraocular part with the electrode array is inserted into the eye via a scleral tunnel and fastened on the retina using a surgical retina tack (titanium tack, GEUDER®, Germany) and corresponding surgical tools. A ring made of medical graded silicone is placed on top of the tack as a retainer. The scleral tunnel is sutured watertight around the implant foil after correct positioning of both implant parts. The conjunctiva is closed over the external part.

**Figure 1:** The retinal stimulator with its epiretinal electrode array on the left side and the encapsuled chip and energy recieving coil on the right.

*Subjects & Methods*

Three patients suffering from RP were enrolled in a prospective study. All patients had a visual acuity of hand-movement or less. Patient´s age ranged from 45 to 65. All patients signed informed consent.

The sugical procedure was performed in two steps. First a preparing surgery including pars plana vitrectomy, cryo- or endolaser retinopexy and, if needed a cataract extraction. The implantation procedure itself followed about 4 weeks after the initial preparing surgery. All surgeries were performed under general anaesthesia. On predefined postoperative dates, stimulation sessions of 3-4 hours were performed using a special stimulation software connected wirelessly to the implanted retinal stimulator. These stimulating sessions included threshold measurements, amplitude modulations as well as motion and pattern perception tests. All results were verified by repeating testing. For threshold testing, the amplitude was increased in predefined steps until the subject reported a phosphene. Point-to-point-discrimination was tested by activation of two consecutive electrodes and the subject was asked, if the phosphene moved from right to left or from top to bottom in a forced-choice interview.

Motion dicrimination was tested by consecutive activation of 4 electrodes in a linear horizontal or vertical direction. The subject was then asked, if the percieved motion was horizontal or vertical in a two-alternative forced-choice interview. Pattern recognition was tested by simultaneous activation of multiple electrodes in a linear horizontal or vertical bar orientation or a cross orientation. The patient was then asked, what kind of pattern he or she percieved in a three-alternative forced-choice interview. Even more complex patterns were tested in a free interview.

Furthermore indocyanine-green-angiographies (ICG, HRA2, Heidelberg Engineering), fluorescein angiographies (Visupac, Carl Zeiss Meditec) and optical coherence tomography (OCT) investigations (Stratus and Cirrus OCT, Carl Zeiss Meditec) were performed on predefined follow-up investigations. The postoperative follow-up was up to 36 months.

The study protocol complied with the principles of Good Clinical Practice and participants were treated in accordance with the tenets of the Declaration of Helsinki. The local ethics committee approved the study.

## 3. Results

*Implantation Procedure*

The preparing surgery was conducted without any notable events in all patients. The preparing retinopexy proved stable and safe retinal conditions for the later implantation. Thus no retinal detachment occured. All implantations were performed without intraoperative complications and all implants were successfully placed epiretinally. The placement of the retinal tack and the fixation of the electrode array on the tack showed no difficulties. In one patient a progressing corneal opacity during surgery made a corneal abrasion necessary to keep the insight sufficient, but the postoperative corneal erosion reepithelialized quickly. Special attention was paid on the suture of the scleral tunnel to avoid postoperative leakage of intraocular fluid and testing for leakage directly after closure showed no percolating fluid in all subjects. Overall the implantation procedure proved feasibility and an easy learning curve for experienced retinal surgeons by using standard vitrectomy tools in combination with only few special and new implantation instruments.

*Stimulating Sessions*

Reported perceptions were predominantly circles, yellow or white in colour. The mean minimum stimulation threshold was 17 nC. The single threshold values showed a high variation according to the subject, the location of the electrode and the date of the stimulation session. An interindividual comparison of threshold values of the same electrode in different subjects showed differences in perception threshold, that was most likely caused by differences of electrode position and distance to the retina [5]. An intraindividual comparison of one electrode in the same patient on different session dates showed a slight change of threshold values even over time, wich sometimes showed a „learining curve“-like lowering of threshold over time, but sometimes also showed higher threshold values on later session

dates. Point-to-point discrimination was successsful in 85% to 89% (right/left, top/bottom). Motion perception was correct in up to 100% in both horizontal and vertical directions. The pattern recognition tests showed correct answers in 98,5% when stimulating a vertical bar, in 89,5% when stimulating a horizontal bar and in 55% when stimulating a cross bar. The lower success rate in correct recognition of the more complex cross pattern goes concordant with the results of the free interview tests of more complex patterns like a triangle or a square (Figure 2). These free answers often showed slight changes from the definite presented pattern, wich is most likely caused by the stimulation of multiple not straight retinotopically organised gangion cells and bipolar cells.

| Pos. | Shape | Describtion | Response |
|---|---|---|---|
| 1 |  | Triangle | 1. a 45 degrees slants, legs of a triangle<br>2. sloped rectangle<br>3. kind of triangle, 45 degree leg with additional a small additional object<br>4. horizontal line with slanted lines<br>5. side view of a roof from a house |
| 2 |  | Cross | 1. 2 angeled lines " / \ "<br>2. Angle<br>3. 2 angeled lines " / \ "<br>4. 2 angeled lines " / \ "<br>5. 2 angeled lines " / \ "<br>6. 2 angeled lines " / \ " |
| 3 |  | 2 vertical, parallel lines | 1. Rectangle, upside<br>2. vertical line with a view slanted lines at one side<br>3. like a vertical oriented comb (horizontal line with some vertical lines)<br>4. like a vertical oriented comb (horizontal line with some vertical lines) |
| 4 |  | Square | 1. a triangle<br>2. a slanted line<br>3. a slanted rectangle<br>4. a cube<br>5. part of a circle, a cone |
| 5 |  | 2 horizontal, parallel lines | 1. a slanted rectangle, 45 degrees lines<br>2. slanted lines with slanted thin lines<br>3. 2 slanted, angled lines (like a roof top)<br>4. a screened square |

**Figure 2:** Examples of free interview answers on complex pattern stimulations

*Clinical Follow-Up*

The ophthalmoscopic examinations as well as the fundus photographies showed a stable and well positioned electrode array over the whole followup period (Figure 3). Only a slight change of retinal vessel configuration near the implant was observed in one subject. There were no signs of tissue proliferation or cell migration onto the implant and electrode array as well as no signs of proliferative vitreoretinopathy. The fluorescein- and ICG-angiographies showed an unchanged and intact blood-retina-barrier in all subjects over the whole followup period. No signs of leakage were detected, especially no signs of choroidal neovascularisation. OCT investigations show the eletrode array as a high refletive thin membrane over the retina. Typical optical shadows can be seen behind the conductant pathways (Figure 4). OCT data proved a close proximity between the retina and the electrode film, while differences in the distance between the electrode array and the retinal surface were not unusual and dependent on the exact location of the OCT scan. Even a slight compression of the retina at the edge of the implant foil could be seen on some OCT scans. In one case OCT scans showed a transient retinal edema. Overall no severe retinal tissue damage was observed.

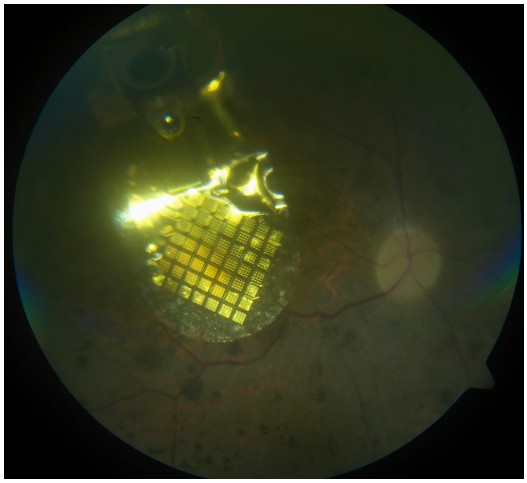


**Figure 3:** Fundus photography 2 years after implantation shows a stable and well positioned implant and electrode array in the macular area.

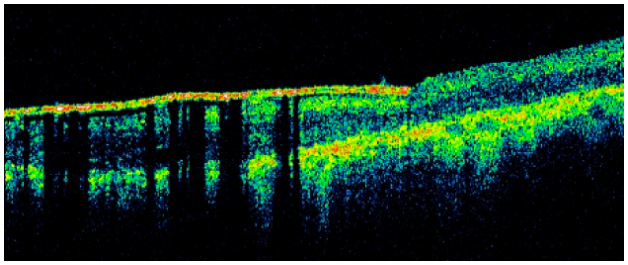


**Figure 4:** OCT of the implant foil. Notice the typical optical shadows of the conductant pathways and the slight compression of the retina on the edga of the implant

## 4. Conclusion

The implantation procedure itself proved feasibility and is easy to learn for the hands of an experienced vitreoretinal surgeon. This was achieved by the use of standard vitrectomy techniques only extended by a few special instruments as the tack positioner or implant guidance forceps. In comparison to subretinal implantation procedures, in wich a retinal detachment needs to be perfomed and a postoperative silicone oil endotamponade is essential after implantation [6], the techniques for epiretinal implantation are easier and faster [7]. Furthermore the wireless data and energy transmission allows a complete wound closure without any cables or connectors penetrating through the wound and increasing the risk of postoperative infections. Nevertheless it is presumed that subretinal implantation can lead to postoperative choroidal neovascularisation (CNV) [8]. We can report no case of CNV in any of our subjects over the whole followup period. These facts point out the safety of the presented epiretinal implant device even during long term implantation and for the intended everyday use. Even more the completely wireless signal and energy transmission to the implant allows unrestricted eye movement and is therfore suitable as a long-term solution.

The activation of the implant and the stimulation sessions are also very encouraging for further implant generations. The threshold values needed to elicit phosphenes are on a safe level and surprisingly much lower than in earlier studies [9]. This might result from an even closer electrode-retina contact by tack fixation in this chronic study compared to a handheld electrode

device in the acute study [4]. We found high success rates in point-to-point and motion recognition as well as in simple patterns similar to letters. This is a crucial point in improving patients´ independence by restoring their ability to read [10,11].
The slightly incorrect answers in the free interview tests with complex patterns were most likely caused by the stimulation of multiple not straight retinotopically organised gangion cells and bipolar cells, as already mentioned above. This phenomenon indicates the necessity of a learning image processor in the pathway between the camera and the electrode array for a visual prosthesis [12]. The captioned image cannot be stimulated directly onto the retina without considering the individual retinotopic conditions of each patient. Otherwise these slight changes in the percieved patterns will result in distortion and complete inability of recognition. Therefore a computed and learning image processor will be implemented in future generations of our implant.
In conclusion the presented results show a safe to implant and effective working wireless epiretinal implant device. The clinical and stimulation data are encouraging for upcoming long-term implantations and the intended every day use via a connected camera. Further improvement is expected with a higher number and efficacy of electrodes and a reduction of the implant´s and especially the electronics´ size [13]. Furthermore a learning image processor in the signal pathway will improve the resolution and recognition rate of more complex images.

## References


[1] Le Roy C: Où l'on rend compte de quelques tentatives que l'on a faites pour guérir plu- sieurs maladies par l'électricité. Hist Acad R Sci (Paris), Mém Math Phys 1755;60:87–95.

[2] Zeitz O, Keserü M, Hornig R et al.: Artificial Sight: Recent Developments. Klin Monatsbl Augenheilkd 2009;226(3):149-53

[3] Feucht M, Laube T, Bornfeld N, et al.: [Development of an epiretinal prosthesis for stimulation of the human retina]. Ophthalmologe 2005;102:688-691.

[4] Hornig R, Laube T, Walter P, et al.: A method and technical equipment for an acute human trial to evaluate retinal implant technology. J Neural Eng 2005;2:S129-134.

[5] de Balthasar C, Patel S, Roy A, et al.: Factors affecting perceptual thresholds in epiretinal prostheses. Invest Ophthalmol Vis Sci 2008;49:2303-2314.

[6] Stingl K, Greppmaier U, Wilhelm B et al.: E.Subretinal visual implants. Klin Monbl Augenheilkd. 2010;227(12):940-5.

[7] Caspi A, Dorn JD, McClure KH, et al.: Feasibility study of a retinal prosthesis: spatial vision with a 16-electrode implant. Arch Ophthalmol. 2009 Apr;127(4):398-401.

[8] Matthaei M, Zeitz O, Keserü M, et al.: Progress in the Development of Vision Prostheses. Ophthalmologica 2011;225:187-192

[9] Humayun M, Propst R, de Juan E, Jr., et al.: Bipolar surface electrical stimulation of the vertebrate retina. Arch Ophthalmol 1994;112:110-116.

[10] Benav H, Bartz-Schmidt KU, Besch D et al.: Restoration of useful vision up to letter recognition capabilities using subretinal microphotodiodes. Conf Proc IEEE Eng Med Biol Soc. 2010;2010:5919-22.

[11] Humayun MS, de Juan E, Jr., Weiland JD, et al. Pattern electrical stimulation of the human retina. Vision Res 1999;39:2569-2576.

[12] Eckmiller R, Neumann D, Baruth O: A tunable retina encoder for retina implants: why and how. Journal of neural engineering 2005;2:9l-104.

[13] Behrend M, Ahuja A, Humayun M et al.: Resolution of the Epiretinal Prosthesis is not Limited by Electrode Size. IEEE Trans Neural Syst Rehabil Eng. 2011 Apr 19.

# Protecting Multicast Session in Survivable WDM Mesh Networks

Baibaswata Mohapatra, Rajendra K. Nagaria and Sudarshan Tiwari, MIEEE
Department of Electronics & Communication Engineering,
M.N. National Institute of Technology, Allahabad-211004, India,
e-mail: {rel0707, rkn, stiwari}@mnnit.ac.in

**ABSTRACT**

Protecting *light-tree* based multicast connections are more demanding than unicast connections in wavelength division multiplexing (WDM) based mesh networks. In this work, we proposed two mathematical models and a heuristic based approach to protect multicast connections in WDM mesh networks. The mathematical models presented in this paper are based on multiple unicast connections. Link-protection and path-path protection schemes are considered to protect the multicast connections. Our heuristic based approach based on sub-path protection technique namely shared-segment protection (SSP) scheme. We simulate the above techniques on conventional networks and the results are compared to evaluate their performances. Simulation results show that SSP scheme offers a reduction in average cost for protecting multicast connections. Static and dynamic multicast connections have been used to calculate the average cost and the blocking probability respectively.

**Keywords:** light-tree, multicasting, protection, survivability, WDM networks.

## 1. INTRODUCTION

WDM optical networks are high speed networks and provide many potential applications [1-5]. In order to support multicast based applications the WDM network be required to employ multicast-capable wavelength-routing switches at the network node [6]. Theses switches are capable to replicate data stream from one input port to multiple output ports. There are two types of switch architectures are used [7]. Opaque switch which support electronic cross-connects with optical-electrical-optical (O-E-O) conversion and transparent switch architecture which support all optical cross-connects (OXCs). Significant work has been carried out to protect unicast connections in WDM networks, however, little effort has been made for protecting multicast connections in WDM mesh Networks.

Several schemes have been proposed in the literature to protect the multicast connections. The simplest idea to protect the multicast tree from single fiber failure is to compute a link disjoint backup tree is reported in [7]. Two trees are said to be link-disjoint if they do not have any common links. The drawback of this scheme is that it uses excessive network resources and also it is not always possible to find a link-disjoint backup tree. Ring based approach is proposed to protect multicast session in [8]. In this approach a ring has to be formed taking the source node and all the destination nodes participating in the multicast session. The limitation of this scheme is that, it is not always feasible to form such ring in all cases, particularly, when the session sizes are large. Hence the scheme leads to higher blocking probability. Segment protection scheme is reported in [9]. The author derived arc-disjoint backup segment for every segment on the primary light-tree. A segment in a multicast tree is defined as the sequence of edges from the source or any splitting node (on a tree) to a leaf node, or, to a downstream splitting node. A destination node is always considered as a segment end node because it is either a leaf node in a tree or a splitting node. The author also proposed optimal path pair based shared disjoint-path (OPP-SDP) scheme. OPP_SDP scheme finds optimal path pair between every source-destination pairs to protect the multicast connection. The author concluded that OPP-SDP is the best scheme w.r.t. average cost and blocking probability among other proposed schemes. Wavelength sharing can significantly reduces the unnecessary capacity reserved for protecting light-trees and they are classified mainly into three categories (self-sharing, intra-request sharing and inter-request sharing) [10-11]. Multicast protection scheme through spanning path is proposed to protect multicast session is reported in [12-14]. A spanning path in a multicast tree is defined as a path from a leaf node to any other leaf node in the light-tree. The scheme derives backup paths for every spanning path in the multicast-tree. The author used the above three kind of sharing and reported as the efficient technique in terms of average cost and blocking probability. Two heuristics based solution is reported for protecting light-trees such as optimal path pair based removing residual links (OPP-RRL) and source leaf path based avoiding residual links (SLP-ARL) in [15]. The author concluded that the scheme is better than OPP-SDP scheme.

In this paper we have extended the ILP based formulation used to protect the unicast (P2P) connections, i.e. a multicast connection is considered as multiple unicast connections from the source node to all the destinations

nodes participating in the multicast session. The rest of this paper is organized as follows. In Section 2, we discuss the proposed methods. In section 3 we present the numerical results and conclusions of this work is presented in section 4.

## 2. PROPOSED METHODS

### A. Problem Description

A multicast connection can be realized as multiple unicast connections. Since in a multicast based application, several replicas of information are routed to the desired destinations, hence a single wavelength channel needs to be assigned for each light-tree.

In mathematical models (link-protection and path-protection), we consider single wavelength channel per fiber where as in heuristic approach (SSP), multiple wavelength channels per fiber are considered.

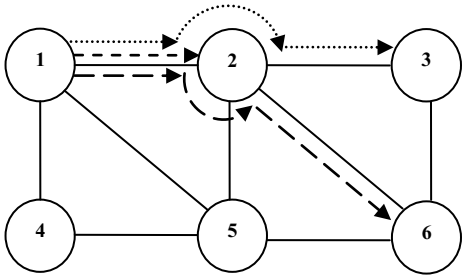
We assume that multicast connections or a light-tree is unidirectional in nature from source to each destination. The session size indicates the number of nodes takes part in a multicast session. In link-protection and path-protection techniques we consider one protected multicast session where as in SSP algorithm, multiple protected multicast connections are considered.

### B. Mathematical Formulation

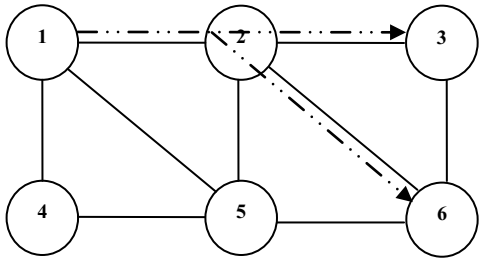
The ILP formulations in [16] are used for finding the minimum cost path to all the destinations. We then explain our approach with a suitable example.

*Example*:

Suppose a multicast connection arrive as S = {1, 2, 3, 6} i.e. node '1' acts as the source and '2', '3' and '6' are the destinations of the multicast session. If we can consider this as three unicast connections then, three different lightpaths of $P_1$= (1,2), $P_2$= (1,2) (2,3) and $P_3$= (1,2) (2,6) needs to be assigned different wavelengths shown in fig. 1 (a). But, if we consider this is a multicast connection then, a single wavelength has to be used to send the information to all the destinations. Hence the light-tree $T = P_1 \cup P_2 \cup P_3$ = (1,2) (2,3) (2,6) could be derived and is shown in fig. 1 (b).



(a)



(b)

Figure 1: (a) Three unicast connections. (b) One multicast connection taking union of three unicast connection shown in fig 1.(a)

**Sets and Parameters:**

*N: Set of nodes*
*L: Set of links*
*E: Set of edges contains arc (i,j) and (j,i)*
*W: Number of wavelength on each fiber link*
*$d_{ij}$: Number of requested connections between node pair i and j*
*D: Set of demand pairs* $(i,j) \in D$ *implies that* $d_{ij} > 0$
*$P_{ij}$: Set of paths between node pair i and j,* $\forall (i,j) \in D$
*R: Set of all paths*
*$A_{ij}$: Set of paths contain arc (i,j)* $\in E$
*$Z_{st}$: Set of all paths from 's' to 't' except the direct link (s,t)*
*$V_{ij}^{st}$: Set of directed paths accessible for restoration from i to j when (s,t) fail.*
*$\tilde{V}_{ij}^{st}$: Set of directed paths not accessible for restoration from i to j when (s,t) fail.*

**Variables:**

*$w_{ij}$: Primary capacity on link (i,j)*
*$s_{ij}$: Spare capacity on link (i,j)*
*$x_p$ : 1 if lightpath p exists, 0 otherwise.*
*$y_p^{st}$: 1 if lightpath exists on path p, when link (s,t) fail, 0 otherwise*.

To establish a primary light-tree the ILP formulation can be written as

$$\text{minimize} \sum_{(i,j)\in L} w_{ij}$$

Demand between node pair: $\forall (i,j) \in D$

$$\sum_{p\in P_{ij}} x_p = d_{ij} \qquad (1)$$

Working link capacity: $\forall (i,j) \in L$

$$\sum_{p\in A_{ij}} x_p + \sum_{p\in A_{ji}} x_p \leq w_{ij} \qquad (2)$$

$$w_{ij} \geq 0 \;\; (i,j) \in L \qquad (3)$$

$$x_p \geq 0 \quad \forall p \in R \qquad (4)$$

Equation (1) ensure that the number of lightpath demand $d_{ij}$ between the node pair (i, j) must be equal to the number of paths '$x_p$ ' within the $P_{ij}$. In a multicast connection this is equal to 1. Equation (2) says about the

number of connection passes in the direction 'i' to 'j' and 'j' to 'i' through the fiber link (i, j). Equations (3) and (4) ensures the non-negative and integer variables. This formulation provides the path from the source to all the destinations with minimum capacity. So taking union of all the selected paths we will get the set of links that a light-tree consists.

**Spare Capacity Allocation Model**

- **Shared Link Protection (SLP)**

In link based protection, we assume that each node is capable to detect a link failure and run a rerouting algorithm around the failed link. The traffic is rerouted only around the failed link. In this strategy, for calculating backup capacity we should have knowledge about the working capacity i.e. the number of primary lightpaths passes through the failed link. The technique followed to form a light-tree for a multicast connection, the backup paths for the failed link is calculated similarly.

TABLE I

| Working link | Backup route |
|---|---|
| (1,2) | (1,5) (5,2) |
| (2,6) | (2,5) (5,6) |
| (2,3) | (2,6) (6,3) |

TABLE I presents the links for the primary light-tree and the backup paths for the corresponding link. Taking the union of the entire backup route gives the cost of backup path which is 5 wavelength links. Since from the above backup capacity we can see that the link (2, 6) is being considered for both primary light-tree and the backup path shown in TABLE I.

So the ILP formulation can be written as

$$\text{minimize} \sum_{(i,j)\in L} w_{ij} + s_{ij}$$

Lost link capacity: $\forall(s,t) \in L$

$$\sum_{p\in Z_{st}} y_p^{st} = w_{st} \qquad (5)$$

Spare link capacity: $\forall(s,t) \in L, (i,j) \in L\backslash(s,t)$

$$\sum_{p\in A_{ij}} y_p^{st} + \sum_{p\in A_{ji}} y_p^{st} \le s_{ij} \qquad (6)$$

Fiber capacity limit: $\forall(i,j) \in L$

$$w_{ij} + s_{ij} \le W \qquad (7)$$

$$s_{ij} \ge 0 \ \ (i,j) \in L \qquad (8)$$

$$y_p^{st} \ge 0 \ \ \forall(s,t) \in L, \qquad \forall p \in R \qquad (9)$$

where $w_{ij}$ may be calculated from the working capacity model. Equation (5) ensures that the number of connections affected due to failure of link (s, t). The affected connections must be rerouted through all the paths from 's' to 't' except the failed link (s, t). Equation (6) ensures that the sum of backup paths passes through link (i, j) and (j, i) must be less than the total number of backup paths available. Equation (7) ensures that the sum of the primary lightpaths and the backup lightpaths on any link is limited to the number of wavelengths available on the fiber link. In our work we have considered homogeneous network i.e. all the fiber links have the same number of wavelengths. Equation (9) and (10) ensures non negative and integer variables.

- **Shared Path Protection (SPP)**

In our second formulation we apply shared path protection technique used for unicast connection.

TABLE II

| Working link | Backup route |
|---|---|
| (1,2) | (1,5) (5,2) |
| (2,6) | (1,5) (5,6) |
| (2,3) | (1,5) (5,6) (6,3) |

Taking union of the entire backup route gives the cost of backup path which is 4 wavelength links and shown in TABLE II.

Hence the ILP formulation can be written as

$$\text{minimize} \sum_{(i,j)\in L} w_{ij} + s_{ij}$$

Spare link capacity: $\forall(s,t) \in L, (i,j) \in L\backslash(s,t)$

$$\sum_{p\in A_{ij}} y_p^{st} + \sum_{p\in A_{ji}} y_p^{st} \le s_{ij} \qquad (10)$$

Lost capacity in forward direction: $\forall(s,t) \in L, \forall(i,j) \in D$

$$\sum_{p\in V_{ij}^{st}} y_p^{st} = \sum_{p\in \widetilde{V}_{ij}^{st}} x_p \qquad (11)$$

Lost capacity in reverse direction: $\forall(s,t) \in L, \forall(i,j) \in D$

$$\sum_{p\in V_{ji}^{st}} y_p^{st} = \sum_{p\in \widetilde{V}_{ji}^{st}} x_p \qquad (12)$$

Fiber capacity limit: $\forall(i,j) \in L$

$$w_{ij} + s_{ij} \le W \qquad (13)$$

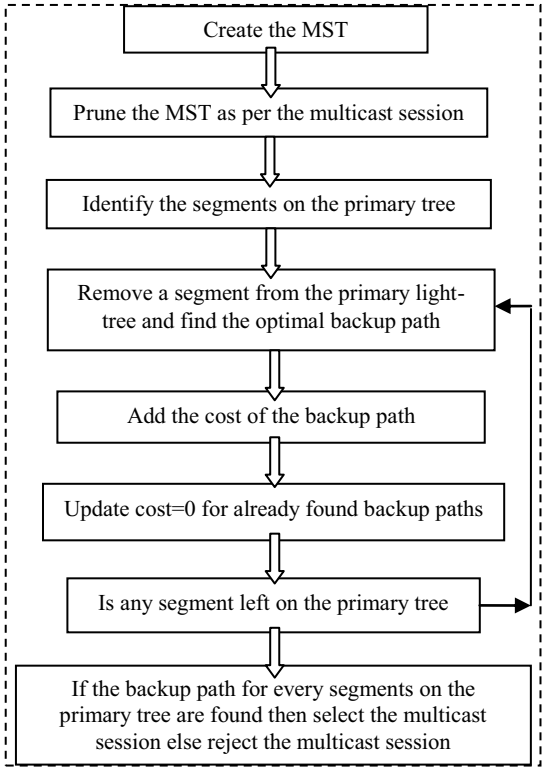
$$s_{ij} \ge 0 \ \ (i,j) \in L \qquad (14)$$

$$y_p^{st} \ge 0 \ \forall(s,t) \in L, \forall p \in R \qquad (15)$$

### C. Heuristic Based Solution

- **Shared Segment Protection (SSP)**

Since the complexity of ILP formulation increases rapidly we study a heuristics approach based on segment protection [9] using minimum spanning tree algorithm such as Prim's. First we generate a multicast session with a fixed group size. The minimum spanning tree (MST) is then pruned with respect to the nodes present in the multicast session. The pruned graph thus formed is the primary light-tree. Then we identify the segments in the primary light-tree. To find the backup paths for each segments in the light-tree remove a segment from the primary light-tree and find the backup (minimum cost path) path to the corresponding segment if exists. If a backup path is found then it is added to backup cost. Update the cost of the path to zero for already found path and repeat the same procedure until all the segments are removed from the primary tree. The backup to the segment is found by Dijkstra algorithm. Source is one end of the segment and destination is the other end of the segment.



Flow model for SSP algorithm

## 3. NUMERICAL RESULTS

We conduct extensive simulations to study the performance of the proposed protection scheme. We compare the proposed schemes with optimal path-pair based shared disjoint path algorithm (OPP_SDP) described in [9], which has been reported as the most efficient scheme by means of average cost and blocking probability.

- **Assumptions and Simulation Setup**

We simulate our schemes on two sample network shown in fig. (2) and fig. (3). For NSFNET the link cost represents the normalized length where as for INDIANET the link cost represents the fiber length between the node pairs in kilometers. We assume that fiber links are bidirectional in nature. The multicast connection is unidirectional in nature from source to destinations. For link-protection scheme and path-protection scheme, we applied precomputed sessions of fixed session size. In the case of SSP and OPP_SDP scheme multicast connections arrive with poisons distribution, and their holding time is set to one unit. The traffic load offered to the sample networks is the number of active multicast session per unit time. We assume that the wavelength converters are not available at the network node. Hence the multicast session or the light-tree occupies single wavelength links throughout their path. As the computational complexity increases with the increase in number of node, we present the average of 200 different multicast sessions for fixed session size for link-protection and for path-protection schemes. On the other hand for SSP and OPP_SDP scheme, we injected $10^5$ different multicast session of same session size. The ILP models are solved with CPLEX 10.2 solver. The SSP and the OPP_SDP algorithm and the k-Shortest path algorithm are implemented using 'C' language.

- **Results and Discussion**

Fig. 4 and fig. 5 shows the average cost for establishing multicast connections versus the session size for NSFNET and INDIANET respectively. Like path-protection and link-protection scheme for unicast connections, the average cost for protecting multicast connection using path-protection technique is much lower than link-protection technique. It is also observed from the above figures that the SSP scheme offers lower cost than the OPP_SDP scheme for protecting multicast session. The author in [9] pointed out that there are cases where backup segments may not be found always, we feel that such case arise rarely. Fig. 6 and fig. 7 shows the blocking probability (BP) offered by SSP and OPP_SDP scheme. It is observed that the BP offered by OPP_SDP is better than SSP scheme till we increase the session size up to approximately 8 and 9 for both NSFNET and INDIANET. But when we increase the session size, it is observed that the BP is almost equal offered by both the scheme as the number of segments increases with the increase in session size. Further we find the BP, keeping the session size in mind; we fix the session size and vary the load. From fig. 8 and fig. 9 it is observed that for NSFNET when we fix the session size equal to 7 the BP offered by OPP_SDP scheme is lower than SSP scheme

where as when we increase the session size to 10 the BP is almost equal. Similarly from fig. 10 and fig. 11 it is observed that for INDIANET when we fix the session size equal to 10 and 15 the BP offered by both SSP and OPP_SDP schemes are almost equal.

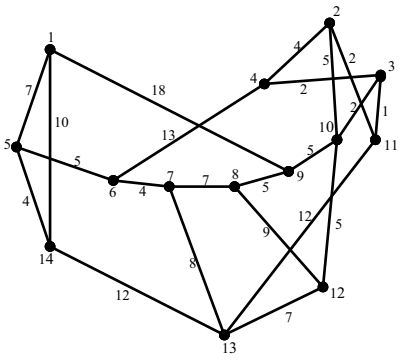


Figure 2: NSF NET (14 nodes, 22 links)

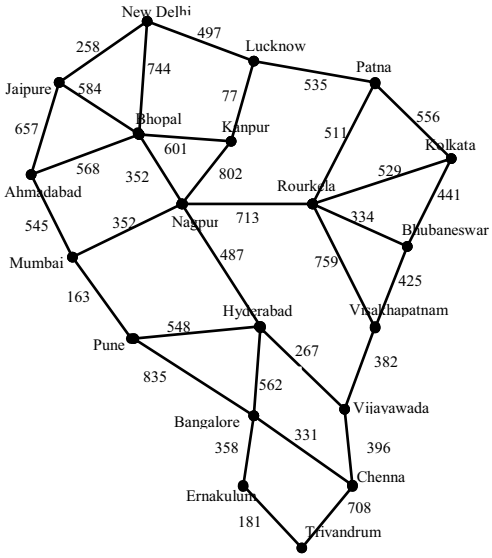



Figure 3: INDIA NET (20 nodes, 33 links)

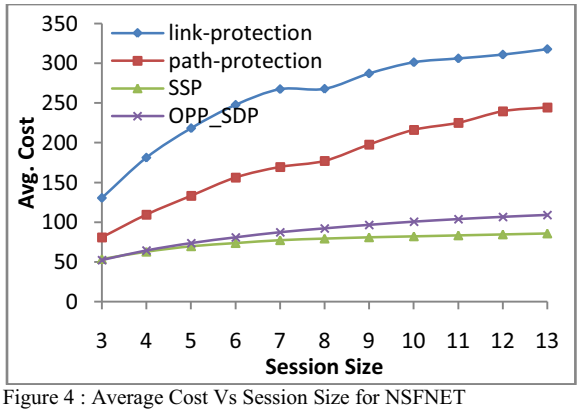



Figure 4 : Average Cost Vs Session Size for NSFNET

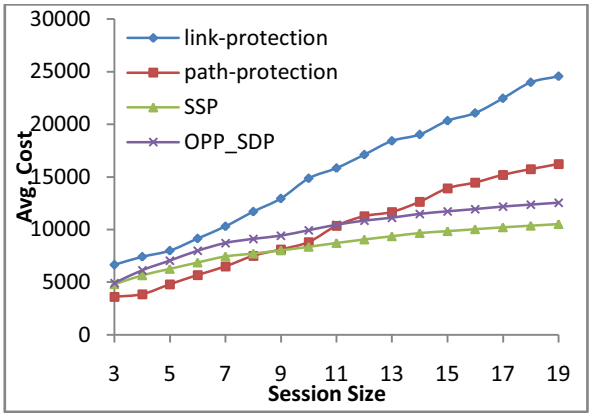



Figure 5: Average Cost Vs Session Size for INDIANET

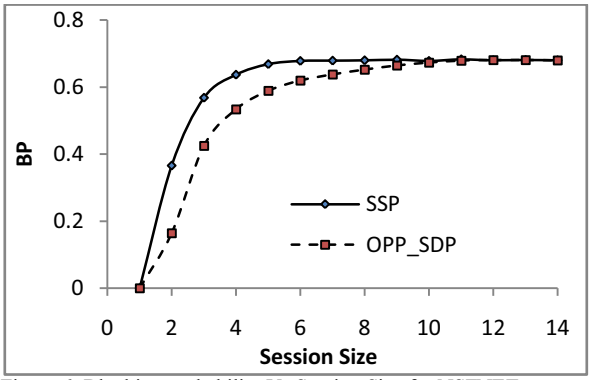



Figure 6: Blocking probability Vs Session Size for NSFNET

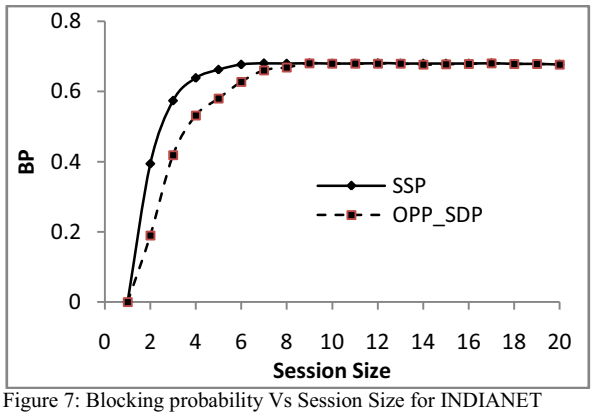



Figure 7: Blocking probability Vs Session Size for INDIANET

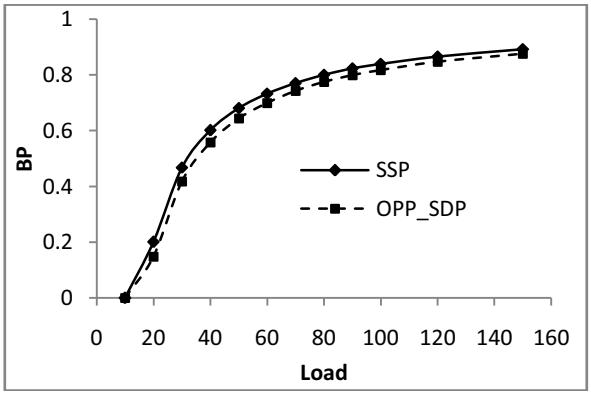



Figure 8: Blocking probability Vs Load with Session Size 7 and 32 wavelengths for NSFNET

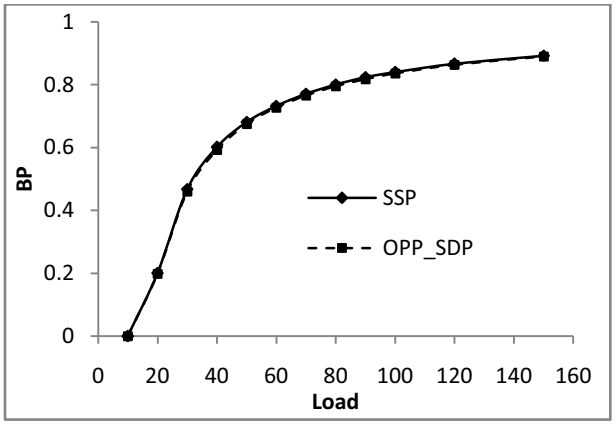


Figure 9: Blocking probability Vs Load with Session Size 10 and 32 wavelengths for NSFNET

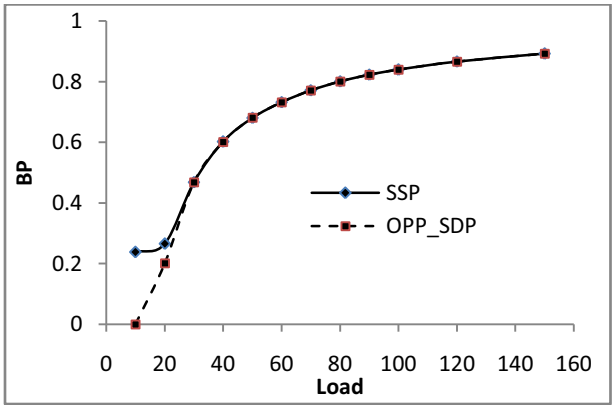


Figure 10: Blocking probability Vs Load with Session Size 10 and 32 wavelengths for INIDIANET

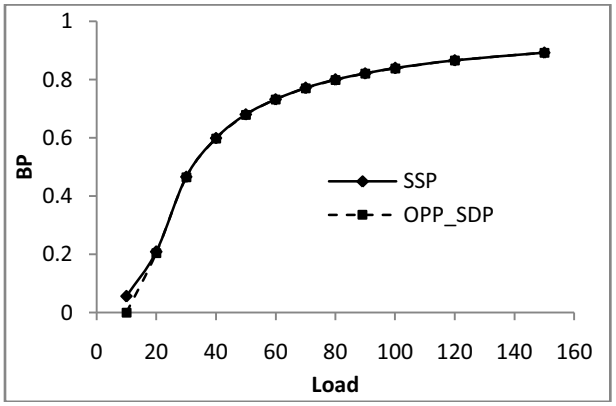


Figure 11: Blocking probability Vs Load with Session Size 15 and 32 wavelengths for INIDIANET

## 4. CONCLUSIONS

In this paper, we have investigated various protection schemes taking static as well as dynamic multicast sessions in WDM mesh networks. The proposed mathematical model for path-protection scheme offered lower cost over link-protection scheme to protect multicast connections. Also the heuristic based approach using shared segment protection scheme offers lower average cost for establishing protected multicast session in the WDM networks. The shared segment protection scheme offer equal blocking probability with the best reported scheme (OPP_SDP), particularly when the session size is larger. Moreover we are investigating further sharing techniques, to design efficient algorithm for protecting dynamic multicast sessions in WDM mesh networks.

ACKNOWLEDGMENT

The first author would like to thank Uday Jalan, student of B. Tech ECE at MNNIT, Allahabad for putting his efforts while implementing the heuristic algorithms.

## REFERENCES

[1]. B. Mukherjee, *Optical Communication Networks*. New York: Mc-Graw-Hill, 1997.
[2]. S. Paul, *Multicasting on the Internet and Its Applications*. Boston, MA: Kluwer, 1998.
[3]. Rajiv Ramaswami and Kumar N. Sivarajan, "*Optical Networks: A Practical Perspective"*, Morgan Kaufmann Publishers, 2nd Edition, 2002.
[4]. C. S. R. Murthy, M. Gurusami, "*WDM Optical Networks - Concepts, Design and Algorithms"*, Pearson Education, Singapore, 2003.
[5]. C. K. Miller, *Multicast Networking and Applications*. Reading, MA: Addison-Wesley, 1999.
[6]. L. H. Sahasrabuddhe and B. Mukherjee, "Light-trees: Optical multicasting for improved performance in wavelength-routed networks," *IEEECommun. Mag.*, vol. 37, Feb. 1999, pp. 67–73.
[7]. N. Singhal and B. Mukherjee, "Protecting multicast sessions in WDM optical mesh networks," *J. Lightwave Technol.*, vol. 21, Apr. 2003, pp. 884–892.
[8]. C. Bowortummarat, L. Wuttisittikulkij, and S. Segkhoonthod, "Lighttree based Protection Strategies for Multicast Traffic in Transport WDM Mesh Networks with Multi-fiber Systems", in *Proc. IEEE ICC'04*, June 2004, vol. 3, pp.1791–1795.
[9]. N. Singhal, L. sahasrabuddhe, and B. Mukherjee, "Provisioning of Survivable Multicast Sessions Against Single Link Failures in Optical WDM Mesh Networks", *IEEE/OSA J. of Lightwave Technology*, vol. 21, no. 11, pp. 2587–2594, Nov. 2003.
[10]. N. Singhal, C. Ou, and B. Mukherjee, "Shared Protection for Multicast Sessions in Mesh Networks", in *Proc. IEEE OFC'05,* pp. 823-825, 2005
[11]. N. Singhal, C. Ou, B. Mukherjee, "Cross-sharing vs. Self-sharing Trees for Protecting Multicast Sessions in Mesh Networks", *J of Computer Networks*, vol 50, no. 2, pp. 200-106, Feb. 2006.
[12]. H. Luo, H. Yu, L. Li, and S. Wang, "On Protecting Dynamic Multicast Sessions in Survivable Mesh WDM Networks", in *Proc. OFC'2006*.
[13]. H. Luo, L. Li, and H. Yu, "Algorithm for Protecting Light-trees in Survivable Mesh Wavelength-division-multiplexing Networks", *J. of Optical Networking*, vol. 5, no. 12, pp. 1071–1083, 2006.
[14]. Hongbin Luo, Lemin Li, Hongfang Yu, and Sheng Wang, "Achieving Shared Protection for Dynamic Multicast Sessions in Survivable Mesh WDM Networks", *IEEE J. of Selected Areas in Comm.*, vol. 25, no. 9, pp. 83-95, Dec. 2007.
[15]. Xiong Wang · Sheng Wang · Lemin Li · Yunji Song, "Achieving Resource Reduction for Protecting Multicast Sessions in WDM Mesh Networks", *J. of Photonic Netw. Comm.*, vol.15, no. 2, pp. 131-140, 2008.
[16]. Jeffery L. Kennington, Eli V. Olinick and Gheorghe Spiride "Basic Mathematical Programming Models for Capacity Allocation in Mesh-based Survivable Networks" *Elsevier Science*, *J.of Management Science*, Vol 35, no.6, December 2007, pp.629-644.

# Village Base Renewable Energy Promotions in Indonesia

**Kamaruddin Abdullah**
**Department of Mechanical Engineering, Darma Persada University**
**East Jakarta, 13450, Indonesia**

## ABSTRACT

One of the effort to solve poverty problem, starting 2007, the government of Indonesia has launched a national Energy Self Sufficient Village (ESSV) program aiming, create the spirit of self help for villagers so that they are capable to fulfil their own basic energy needs for cooking, lighting, and to improve their wellbeing by cleverly utilizing locally available renewable energy sources to generate economic activities. By 2010 there were already more than 630 ESSVs have been established throughout the country. Each ESSV establishment has required between IDR 400 million (US$ 47,000) to IDR 1,2 billion (US$141,000) for each ESSV depending of types of renewable energy sources available in the village. The program however, needs continuous long term monitoring and evaluation using university program to ensure it sustainability. The paper describes how Darma Persada University could help mitigating the current problems faced in implementing the ESSV program.

**Keywords**: ESSV program, renewable energy, productive uses, lesson learnt, university roles

## 1. INTRODUCTION

As a concrete effort to solve poverty problem and energy crisis, in 2007, the government of Indonesia has launched a national program called the Energy Self Sufficient Village (ESSV) [1]. The program aims to create the spirit of self help among the villagers so that they are capable to fulfil their own basic energy needs for cooking, lighting, and to generate economic activities through utilizing locally available renewable energy sources. Through the program, implemented by eight related ministries, access to badly needed electricity and fuel for the rural poor has been made available by providing micro-hydro, solar PV, biogas, and bio-fuel production infrastructures. By 2010 there were already more than 630 ESSVs have been established. Each established ESSV required between. It is envisaged that by the year 2014, cumulatively 3000 villages is expected be established by the program. Such efforts by the government in fact are in line with Darma Persada University's development technology concept of Energy, Economy and Environment ($E^{3i}$ ) village also launched in 2007.[2].

Despite of this good intention and initiative by the government to provide the village with easy access to energy, however, at present, the villagers are, only capable to utilize the provided energy facilities for household cooking fuel and lighting. Under this condition the potential electricity generated have been kept mostly idle during the day time. A study is now underway in two ESSV samples to determine the impact of renewable energy intervention on sustainable economic development of the selected villages. For this purpose survey to two sample ESSVs, namely, Gunung Halu village in West Java, and Banyumeneng village in Special Province of Jogjakarta. LEAP (Long term Energy Alternative Planning System) model and I/O Table will be used to study the impact of renewable energy technology intervention, on sustainable development of sample villages. Hybrid solar dryer is proposed to substitute direct sun drying practices in both villages. The LEAP model will also be applied to study how far the existing 18 kWe, micro-hydro at Gunung Halu village and the existing 1200 Wp solar PV water pumping can change the economic structure of the sample village by developing new industries based on local resources.[3].

This paper describes RD/D efforts in developing renewable energy based technologies which later could be utilized for value added activities in the two sample ESSVs.

## 2. RENEWABLE ENERGY POTENTIAL AND DEVELOPMENT ROADMAP

### 2.1. Renewable energy potential

Table 1, shows the renewable energy resources potential as well as the condition of fossil fuel in Indonesia.[4]. As shown, geothermal, hydro and biomass wastes, are among the largest energy resources in the country. Biomass energy resources from oil bearing plants such as *jatropha curcas*, oil palm and other ingenious plant have the great potential to substitute partially diesel oil, while cassava, corn, sugar cane and *arenga* palm can be extracted to produce bio-ethanol as gasoline blend.

Table 1. Renewable and fossil fuel resources of Indonesia

| NO | NON FOSSIL ENERGY | RESOURCES (SD) | INSTALLED CAPACITY (KT) | RATIO KT/SD (%) |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 = 4/3 |
| 1 | Hydro | 75,670 MW | 4,200 MW | 5.55 |
| 2 | Geothermal | 28,53 GW | 1,189 MW | 4.2 |
| 3 | Mini/Micro Hydro | 500 MW | 86.1 MW | 17.56 |
| 4 | Biomass | 49,810 MW | 445 MW | 0.89 |
| 5 | Solar Energy | 4.80 kWh/m$^2$/day | 14.1 MW | - |
| 6 | Wind Energy | 3 – 6 m/s | 1.4 MW | 0.015 |
| 7 | Uranium | 3.000 MW (e.q. 24,112 ton) for 11 years*) | 30 MW | 1.00 |

*) only in Kalan – West Kalimantan

| No | FOSSIL ENERGY | RESOURCES (SD) | RESERVES (CAD) | RATIO SD/CAD (%) | PRODUCTION (PROD) | RASIO CAD/PROD (YEAR)*) |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 = 4/3 | 6 | 7 = 4/6 |
| 1 | Oil (billion barrel) | 56.6 | 7.99 **) | 14 | 0.346 | 23 |
| 2 | Gas (TSCF) | 334.5 | 159.64 | 51 | 2.9 | 55 |
| 3 | Coal (billion ton) | 104.8 | 20.98 | 18 | 0.254 | 83 |
| 4 | Coal Bed Methane/CBM (TSCF) | 453 | - | - | - | - |

*) With assumption no new resource discovered

Solar energy is available all over the country at 4,8 – 5,2 kWh/m$^2$/day, and can be used as distributed power supply such as for Solar Home System, but can be integrated with other energy sources such as wind and biomass to operate Small Processing Unit in the village. Wind is available in limited locations in the country, where wind speed is available between 3 – 6 m/s.

In order to anticipate the energy crisis in 2005 the government had enacted laws and regulations to promote effective utilization of the potential renewable energy sources nation wide. Presidential Regulation N0 5, 2006, for example, stipulated the national energy policy, implementation of BPPEN 2005-2025 (National Energy Management Blue Print), and in 2007 the Energy Law No.30, 2007 was enacted as an "umbrella" to all related energy regulations issued so far.

### 2.2. Renewable energy road map

Based on these laws and regulations, national energy road map was issued, including the critical role to be born to the development new and renewable energy sources, which based on the energy conservation scenario, Indonesia will need around 3000 MBOE (Million of Barrel of Oil Equivalent). In which initially the share of new and renewable energy was around 17%. (renewable energy sources, around 13%) by the year 2025 Later with the establishment of the Directorate General of New ,Renewable Energy and Energy Conservation (NREEC) in 2010, the share of new and renewable energy sources was increased to 25% (See Fig.1).

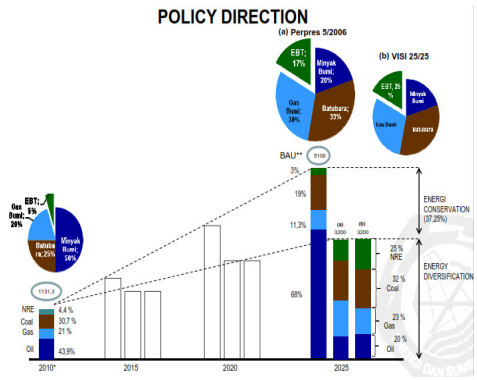



Figure 1. New and renewable energy development road map [2].

## 3. THE ENERGY SELF-SUFFICIENT VILLAGE PROGRAM

One of the government effort to accelerate the utilization of the potential renewable energy sources of the country, in 2007, the government has launched a national program called the Energy Self Sufficient Village (ESSV), implemented by eight related ministries under the coordination of the Coordinating Minister of Economy [1]. The purpose of this program was to establish the spirit of self help among the less developed rural areas with locally available renewable energy sources, namely, bio-fuel, micro-hydro, solar, wind and biomass energy ( other than biodiesel and bio-ethanol). By providing the needed energy access to around 68% of rural population of the country it was expected that economic growth can be accelerated in rural areas, creating new job opportunities and eventually improve the quality of village life. Since 2007, the government had invested between IDR 400 million (US$ 47,000) to IDR 1,2 billion (US$141,000) for each ESSV, depending of types of renewable energy sources available in the village mostly for establishing basic renewable energy generating facilities. In implementing the program, the government have been persuading all stake holders, such as local government, state enterprises, including the private sectors not It was reported that there are already more than 630 ESSVs had been established under the program. In the latter year of ESSV project implementation the government has also begun to provide the ESSV with productive activities such as for coffee processing, production of bio-diesel and bio-ethanol, rice milling, cocoa processing, etc. It was intended that after three years the project will be handed over to local government to continue central government initiatives until sustainable condition can be achieved. Survey to some of the ESSVs had indicated that most of the visited villages were still capable to utilize the generated energy to fulfil their need for cooking and lighting only, very few were capable to utilize the generated energy for productive uses. Consequently, micro-hydro with 20 to 100 kWe capacity installed in some ESSVs have been keep idle and wasted during day time. The survey also indicated that the ESSV will need long term support both in terms of financial and technical aspect on how to utilize the provided energy facilities for their outmost benefit for economic wellbeing.

In summary, the ESSV program has successful in providing access to basic need of energy for the rural people but still need further improvement in the following aspects:

1) Provision of long term financial support ( 5 -10 years may be necessary) such as easy access to innovative financing scheme to ensure sustainability of the program
2) Continuous guidance through effective monitoring and evaluation activities, to improve managerial skill of the beneficiaries by providing access to working capital, better production method to ensure good quality of processed products and find captive market of their product
3) Long term energy planning skill such as using LEAP software and I/O Table so that the provided energy access can induce the growth of industries and employment creation activities.

## 4. ROLE OF UNIVERSITY

University can play role by undertaking the ESSV program as part of their academic program The study may relate to social engineering on how local villagers can participate in the development programs, develop and apply innovative financing scheme and management institution skill of the beneficiaries of the facility installed, including providing guidance to create market of their processed products. The university may also conduct renewable energy based RD/D technologies relevant to felt need of the local people. Such an attempt has been continuing until today, and several innovations had been introduced and field tested. Figs.2. and 3 show typical examples of RD/D results on solar drying applications and Fig.4, shows how hybrid nocturnal cooling unit can help to extend shelf life of fruits and vegetables in the village at competitive costs as compared to the conventional storage system. Each of this developed individual technology can be integrated to form a Small Processing Unit (SPU), shown in Fig.5, in

which staple food such as rice, including corn and soybean and high value crops such as coffee, cacao and peppers can be processed into finished products in the form of ground coffee, cocoa pastes, milled rice, corn, etc.

During a national seminar and training of productive uses of locally available renewable energy sources in March 2007, Darma Persada University, has introduced a development technology concept called the $E^{3i}$ (independent in Energy, Economy and Environment) village, aiming to establish villages capable to utilize cleverly their potential renewable energy sources for economic activities keeping harmony with the environment.[2]. This concept can be regarded as complimentary to the government ESSV program since the former has been concentrated on directly providing Small Processing Unit (SPU) in the villages powered by available renewable energy resources.

It was recommended that in order to ensure the sustainability of the undertaking, university programs such as annual field practices of undergraduate students, under graduate and postgraduate research works as partial fulfilment of their graduate thesis or dissertation.

Figure 2. Hybrid recirculation type solar ICDC under trial test.

Figure 3. Hybrid ICDC solar dryer for leather industry

Figure 4. Hybrid nocturnal cooling for temporary cool storage in the highland of Malang, East Java[5].

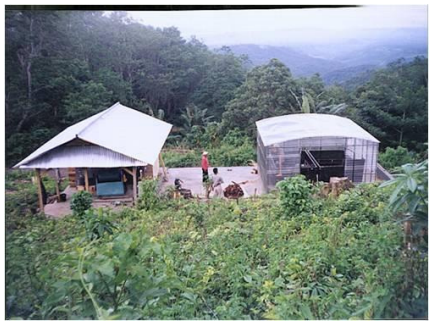


Figure 5. Small Processing Unit model installed by CREATA-IPB in 1999 in Sumbawa Island. [5]

## 5. SUSTAINABLE PARAMETERS

### 5.1. The production function.

Sustainable parameter proposed here can become an effective tool in evaluating the ongoing ESSV program. By comparing the parameters obtained from established ESSVs on regular basis improvement in management as well technical aspects of applied renewable energy technologies can then be made to ensure sustainable operation of the provided facilities.

Assume that a renewable energy technology has been installed in a rural cooperative as the main source of energy for a SPU and to be managed by local entrepreneur. The SPU can produce various kind of value added products ready to be sold in the market. Let the production function of the cooperative using any kind of renewable energy technology such as solar dryer mentioned above, can be represented by [3],[6].

$$dC/dt - a\ \ C = f(b) \qquad \text{.....(1)}$$

with IC; at t=0, C=C0

Here C is the number of sales or production per unit time, $a$ is the managerial coefficient (1/unit time) , $b$ is a coefficient related to the effectiveness of the appllied renewable energy technologies (IDR/ unit time) and t is unit time.

Now let f(b)=b, then the solution of eq.(1) becomes, after applying the initial condition,

$$C(t) = C_o e^{at} + \frac{b}{a}(e^{at} - 1) \qquad \text{.....(2)}$$

The value of $a$ and $b$ then can be determined after solving eq..(2) using at least two data, or if more data are available the least square method can be applied.

A survey had been made in ten SSEVs in West and Central Java and in South Sulawesi. Data obtained from sample ESSV with micro-hydro was calculated from the regular payment of electricity used for lighting only since the record for productive uses were not available. The data than was extrapolated in order to generate a production function and for the calculation of the sustainable parameters $a$ and $b$, respectively. Fig. 6 shows the comparison between the production of a micro-hydro utilization in G. Halu, West Java, as compared to eq. (2). The sustainability parameters

obtained were *a*= 0,26 and *b*=0,13, respectively. Fig.7 shows another case where solar drying technology applied for small tanneries to dry tanned leather [7] and Fig. 8. showing sustainable parameters for a hybrid solar-wind and diesel project [8]. All quoted sales or production data in the above examples were estimated by extending the field data obtained during the survey assuming that the SPU could be operated in sustainable manner. True data should be obtained on regular basis and such work can be accomplished by students from university during their field practices or as research materials for their graduation thesis.

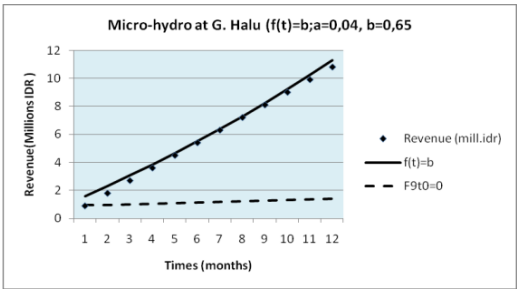



Figure 6. Comparison between data and computed results using eq.(2)

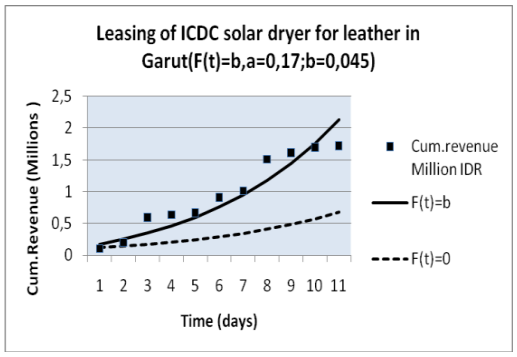



Figure 7. Cumulative leasing data (black marker) and calculated value for f(b)=b, (solid lines) and the case of no technology intervention (broken line).

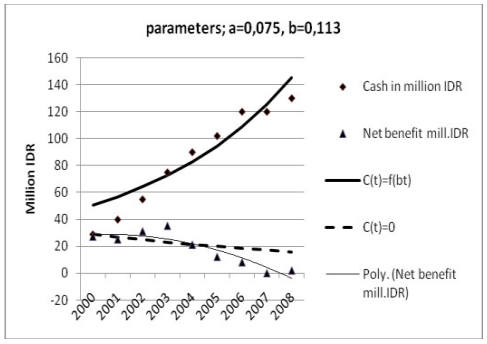



Figure 8. Sustainable parameters for Oeledo Project

## 6. CONCLUSIONS

Village base renewable energy promotion is now taking place in Indonesia. It has been successful at least in providing access to basic energy need for household cooling and lighting, including the possibility to induce economic activities in the village. Nevertheless, direct involvement of university is an imperative action to ensure sustainable operations of the provided energy generating facilities. University involvement may include short and long term energy planning at village level, helping the beneficiaries in implementing priority programs according to the energy plan, help to obtain working capital, provide training on good production practices and to help in finding market of their products. University progma which can be designed to apply holistic and multidisciplinery approached can help on how to manage the facilities properly, so that it could bring a real benefit to the user. Successful undertaking of such university involvement can then be used as working model to other ongoing ESSV programs. A method to determine sustainability of ESSV program has also been introduced in this paper and some results of its application have been presented.

## ACKNOWLEDGMENT

Part of the study have been obtained from research grants conducted under the Graduate Team Research Grant of the Directorate General of Higher Education under contract No:317/SP3/PP/DP2M/II/2006 and from Competence Research Grant, contract No.: 377/SP2H/PLK/Dit Litbangmas/IV/2011.

## REFERENCES

[1].Coordinating Ministry of Economy,RENSTRA Rencana Strategis 2009-2014, Desa Mandiri Energi (ESSV Strategic Plan 2009-2014), 2008..

[2].Kamaruddin A, Establishment of Self Sufficient Village by Utilizating Locally Available Renewable Energy Sources. Proc.National Seminar and Training on Establishment of Self-sufficient Villages Utilizing Locally Available Renewable Energy Sources) , Jakarta, May 22-25, 2007.

[3]. Kamaruddin A. **E3i Self Sustainable Eco-village concept**, UNESCO-E-learning material,2010.

[4]. Directorate General of New and Renewable Energy and Energy Conservation,: Alternative Fuels and the Environment, Paper Presented in a Talk Show on Implementation of Sustainable Energy Policy, Jakarta September 27,2010.

[5]. Kamaruddin A.**Teknologi Berbasis Sumber Energi Terbarukan Untuk Pertanian (Renewable Energy Based Technologies for Agriculture**), CREATA-IPB. Pub. (ISBN, 978-979-96105-4-6), 2007..

[6] Kamaruddin Abdullah, Sustainable Parameters, In Introducing Renewable Energy Technology. **ISESCO Science and Technology Vision**, Vol.5, No.8, 2009,pp.7-10.

[7]. Darma Persada University.Hybrid GHE Solar Dryer for Tanned Leather. Final Report to USAID-SENADA Business Innovation Fund (BIF) project October 2008-April 2009.

[8]. Dauselt, Clause Jürgen, Presentasi Bogor TT Case 2 PV Hibrida IND 030809, www.scribd.com/.../Presentasi-Bogor-TT-Case-2-PV-Hibrida-IND 030809 . visited April 4, 2011.

# Climate policy impacts on energy system: a long-term analysis with the TIMES Integrated Assessment Model (TIAM-FR)

Sandrine SELOSSE
MINES ParisTech, Center for Applied Mathematics
Sophia Antipolis, France

Edi ASSOUMOU
MINES ParisTech, Center for Applied Mathematics
Sophia Antipolis, France

Nadia MAÏZI
MINES ParisTech, Center for Applied Mathematics
Sophia Antipolis, France

## ABSTRACT

This paper analyzes technological options to mitigate $CO_2$ emissions and reach the 2°C objective expressed to UNFCCC since COP15, through global and regional climate objectives. The first aim of this paper is evaluate the impact on global $CO_2$ mitigation possibilities according to carbon constraints scenarios and the level of development of different regions An important result consists in the necessary participation of developing countries in the international climate policy. The second main aim of the study is to study the evolution of the energy system for developed and developing countries. While CCS technologies deployment appears as a necessary solution for fast developing countries (especially China), developing countries (and in a lower manner developed countries) need to invest in renewable energies. But in both cases, countries face economical aspects of their energy transition, such as costs, large scale deployment, traditional energy sector core activities and objectives.

**Keywords**: Energy system, Long-term modelling, TIAM-FR, Climate policy, CCS technologies, Renewables energies and technological choices

## 1. INTRODUCTION

Strategies for reaching the long-term collective goal of limiting global average temperatures increases to 2°C above pre-industrial levels requires a transformation of the world energy system. In this paper, different pathways are analyzed for to achieving a low-carbon economy and climate goals expressed in the Copenhagen Accord.

The evolution of the energy system differs between regions, and high differences in the deployment of (new) cleaner technologies can be observed. Utilizing a global energy system model, we evaluate emission policies in terms of their impact on global and regional energy systems, and their effectiveness to reduce emission and achieve global warming objectives. The model is furthermore used to analyze technological and economic plausibility of strategies deployed to reduce GHG emissions.

## 2. TIAM-FR PRESENTATION

The analyses carried out in this study are based on the TIAM-FR (the French version of the TIMES Integrated Assessment Model) developed under the Energy Technology Systems Analysis Program (ETSAP) of IEA. TIAM-FR is a technology-

rich, bottom-up energy system model. The model depicts the world energy system with a detailed description of different energy forms, resources, processes/technologies and end-uses. Links between commodities and technologies are described via a Reference Energy System, and thousands of technologies are used to model the different sectors of the energy system (energy procurement, conversion, processing, transmission, and end-uses). The model jointly considers the extraction, transformation, distribution, end-uses, and trade of various energy forms and materials.

TIAM-FR is a geographically integrated model in which the world is represented using 15 regions on the time horizon from 2005 to 2100. This study investigates changes in the energy system until 2050. The model aims to supply energy services at minimum global cost by simultaneously making decisions on equipment investment, equipment operation, primary energy supply, and energy trade. Costs considered in the model includes investment costs, operation & maintenance costs, costs of imported fuels, incomes of exported fuels, the residual value of technologies at the end of the horizon, etc.

The main outputs of the model are future investments and activities of technologies for each time period. Furthermore, the structure of the energy system is given as an output, i.e. type and capacity of the energy technologies, energy consumption by fuel, emissions, energy trade flows between regions, transport capacities, a detailed energy system costs, and marginal costs of environmental measures as GHG reduction targets. Indeed, the model considers GHG emissions from fuel combustion and processes, carbon capture and sequestration technologies, and mitigation technological options for $CH_4$ and $N_2O$. The model can thereby be used to analyze atmospheric GHG concentrations and temperature changes [1].

## 3. SPECIFICATION OF SCENARIOS

The development of the energy system was evaluated according two environmental scenarios and a baseline business as usual (BAU) scenario without any emission constraints. Furthermore, the impact of environmental scenarios was analyzed in terms of changes of the global and regional energy systems.

The **Reg_Targ** regional scenario represents the lower $CO_2$ mitigation targets by 2020, as expressed in the Copenhagen January 2010 Agreement by Europe, the United States of America, Australia, Canada, Japan, China and India. Targets by 2050 were assumed according to international convergence in term of $CO_2$ emissions reductions, literature, or expressed political ambitions. Note that the scenario is regional in that it assumes that only the previously described countries are bounded by the climate constraint. These targets, and the corresponding level of $CO_2$ emissions, are expressed in the Table 1.

***Table 1: Cop 15 targets for $CO_2$ emissions***

| Regions | Year ref. | Year target | Targets | Reduc. type | Level CO2 (Gt) |
|---|---|---|---|---|---|
| WEU-EEU | 1990 | 2020 | 20% | Emi. Reduc. | 3,54 |
| | | 2050 | 80% | | 0,89 |
| USA | 2005 | 2020 | 17% | Emi. Reduc. | 4,74 |
| | | 2050 | 83% | | 0,97 |
| AUS | 2000 | 2020 | 5% | Emi. Reduc. | 0,37 |
| | | 2050 | 80% | | 0,08 |
| CAN | 2005 | 2020 | 17% | Emi. Reduc. | 0,47 |
| | | 2050 | 83% | | 0,1 |
| JPN | 1990 | 2020 | 25% | Emi. Reduc. | 0,79 |
| | | 2050 | 80% | | 0,21 |
| CHI | 2005 | 2020 | 40% | CO2 intensity | 6,45 |
| | | 2050 | 10% | Emi. Reduc. | 4,12 |
| IND | 2005 | 2020 | 20% | CO2 intensity | 2,16 |
| | | 2050 | 10% | Emi. Reduc. | 0,98 |

Intermediate targets are also introduced for the USA and Canada regarding their pledges to UNFCCC: 30% for 2025 and 42% for 2030.

The **Glob_50,** global scenario assumes that world $CO_2$ emission in 2050 should be 50% less than that of the 2000 level. The scenario is thus in line with the consensual 2°C objective expressed at COP15 (IPCC, AR4). All regions are bounded by the climate constraint but they are not constraint at a beforehand determinate level of $CO_2$ emissions. Regarding this carbon constraint, world emissions of $CO_2$ could not be higher than 11 Gt.

## 4. RESULTS

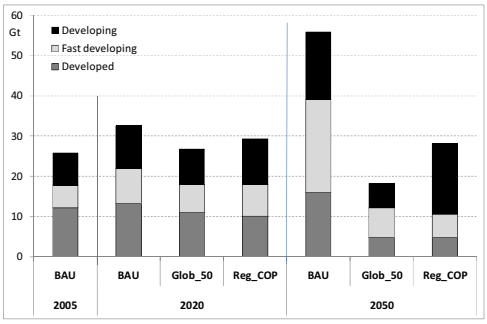
The analysis of the results focuses on the impact of the climate change scenarios on:

- $CO_2$ emissions levels by regions and sectors.
- Electric mix: the use of fossil fuels and the future technological investments (renewables, carbon capture and storage).

**Environmental objectives**

Graph 1 presents $CO_2$ emissions trajectories for the different scenarios and regions. The **Glob_50** and **Reg_Targ**, scenarios, respectively results in a decrease of $CO_2$ emissions by 2050, of 38 Gt and 28 Gt in comparison to the BAU scenario. Thus, the two emission scenarios resulted in highly different levels of emission reductions.

***Graph 1: $CO_2$ emissions by region and scenario***



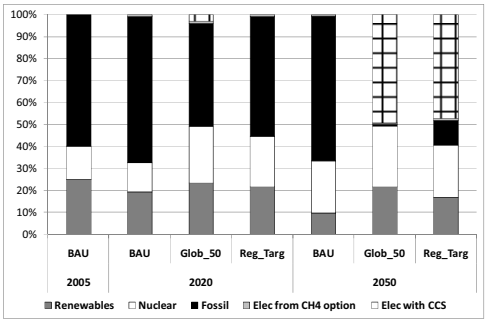

The level of $CO_2$ emissions of developed countries in 2050 was quite the same in the two climate scenario, i.e. around 5 Gt of $CO_2$. It is higher in 2020 in the global scenario, 11Gt of emitted $CO_2$ against 10 Gt of $CO_2$ emitted in the target scenario. This let us think that developed countries are engaged in coherent pathways to reach the global climate ambition of 2°C. The question will be their capability to reach their environmental ambition. This is also the case for China and India with a 2020 and 2050 contribution to the $CO_2$ goal more or less equal for the regional and global scenarios. This confirms an important point of the global climate policy: the world could not reach an acceptable level of $CO_2$ emission without the participation of developing countries [2] [3].

**Technological challenges**

CCS technologies were found to be highly deployed for reaching the climate targets (see graph 2). CCS was found to be particularly important to response to carbon constraint for fastest developing countries, like China, whatever the scenario.

The environmental constraints also lead to an increase deployment of renewable technologies, especially in developing countries in the global scenario where there are constrained for their $CO_2$ emissions.

***Graph 2: World electricity mix***



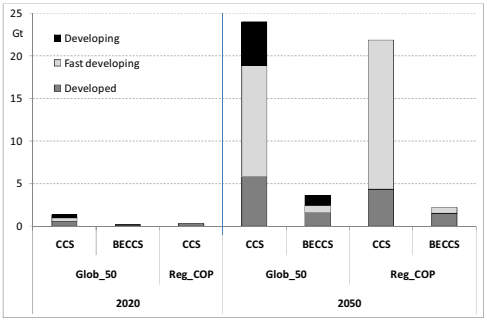

Developed countries also invest in renewables technologies to reach climate targets (whatever the global policy or their COP 15 commitments). In China and India, the deployment of renewables is slowed by investment in CCS technologies. Note that carbon constraint involves an increase of nuclear power in developed and developing countries.

**Carbon Capture and Storage technologies deployment**

According to industry leaders such as Alstom Power, China is expected to be a major market for CCS technologies [4]. Graph 3 highlights the same CCS as an important market opportunity. In the global environmental scenario, 16 Gt of carbon

need to be sequestrated in 2050 if the power generation comes from coal plants.

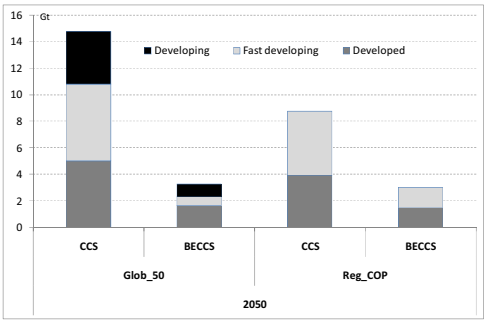
***Graph 3: CCS and BECCS deployment***




However, this ignores the highly development costs to make CCS available at large scale of commercialization and unresolved safety problem. Obtain the social acceptability will be a challenge for the large scale commercialization of CCS technologies.

To highlight this point, a constraint was constructed regarding deployment of coal in China. The constraint was constructed in line with Chinese energy policy to accompany its environmental target. In this scenario, the share of coal in the electric mix from fossil was limited to no more than 50%. The constraint for China was incorporated into the **Glob_50** and **Reg_Targ** scenarios, but note that developed and developing countries are not directly impacted by this change in the constraints.

***Graph 4: CCS and BECCS deployment***




Results according to the climate scenario highlights the reduced investment in CCS technologies to reach climate policy for fast developing countries, and here, more specifically China (see graph 3). With this new constraint, China sequestrates 4.2Gt of $CO_2$ in the global scenario in 2050. However, as a consequence of the constraint, an increased development of renewables energies to generate electricity was noted. However, the deployment of renewable energies may be limited in China, thus forcing a high deployment rate of CCS in China [5].

## 5. CONCLUDING REMARKS

Environmental stakes requires global constraints and $CO_2$ mitigations are required from developed countries and developing countries (including fast developing countries like China and India) for global climate constraints to be reached. Post-COP15 policies thus require $CO_2$ mitigations from developed countries and developing countries. The question is to determine the possible and fair level of $CO_2$ mitigation for regions [4], considering that climate strategies impact on impacts energy systems, and that the action potential differs between regions.

For each region, costs and the most promising technologies to reach climate targets are different. Regional differences exists in terms of the existing energy system, infrastructure, technological potential (renewable energies, CCS, etc.), future economic development, and priorities.

For addressing the problem of global climate change, CCS technologies are expected to be deployed. However, are the required large scale investments in CCS technologies feasible, and can they be deployed to the considered levels? In order to face stringent carbon constraints, CCS technologies need to be installed on an industrial scale, but potential, effectiveness and cost aspects are still highly uncertain. Large scale adoption

of this technology is not yet economically feasible, and commercialization is likely to take time due to high costs and safety problem (leak back into the atmosphere and social acceptability).

This study also highlights that even if international cooperation is needed to face the energy-climate problem, it is not only countries that must act, but technological progress must also be realized.

The pool of existent and future technologies need to be extended and the mitigation potential of the technologies need to be realized. This not only concerns CCS technologies, but also non-fossil technologies, like wind, solar, biomass, etc. Like for CCS technologies, the potential of deployment of renewables is still uncertain. Notably, in the case of intermittent renewables power and their variability, the question remains still remains if a power system can introduce a larger share of renewables without any changes whatsoever? Deployment of renewables energies face traditional energy sector objectives, such as reliable electricity supply [6]. The electricity network will need to be reinforced to ensure stability, thus adding to the cost of renewables. Energy-climate goals to be addressed are reduction in energy-related greenhouse gas emissions (and environment protection more widely), energy security, energy service cost minimization (for both business and residential consumers), and energy service accessibility [6]. These objectives imply not only behavioral changes in consumption (changes we have not evocated here because of a technological focus), but also further technological development. Investments will be required to make a low carbon society feasible.

## 6. ACKNOWLEDGMENT

This research was supported by the Chair Modeling for sustainable development, driven by MINES ParisTech, Ecole des Ponts ParisTech, AgroParisTech, and ParisTech, supported by ADEME, EDF, RENAULT, SCHNEIDER ELECTRIC and TOTAL.

## 7. REFERENCES

[1] R. Loulou and M. Labriet, "ETSAP-TIAM: the TIMES Integrated Assessment Model, Part I: Model structure", **Computational Management Science**, 2007, doi: 10.1007/s10287-007-0046-z.

[2] F. Rong, ''Understanding developing country stances on post-2012 climate change negotiations: Comparative analysis of Brazil, China, India, Mexico, and South Africa'', **Energy Policy**, Vol. 38, 2010, pp. 4582-4591.

[3] F. Gemenne, **Géopolitique du changement climatique**, Paris: Armand Colin, 2010.

[4] M. den Elzen, P. Lucas, D. van Vuuren, ''Abatement costs of post-Kyoto climate regimes'', **Energy Policy, Vol.** 33, 2005, pp. 2138-2151.

[5] L. Lan, "China to take lead in utilizing green technology", **China Daily**, 2010, *http://www.chinadaily.com.cn/business/2010-06/17/content_9982315.htm*

[6] M. Bazilian, H. Outhred, A. Miller and M. Kimble, "Opinion: An energy policy approach to climate change", **Energy for Sustainable Development**, 2010, Vol. 14, No. 4, December 2010, pp. 253-255.

# Supply Curves of Ancillary Services for Planning

**Martin Beneš**
**Dpt. of Energy Economics, Czech Technical University in Prague**
**Prague, Czech Republic**

**Oldřich Starý**
**Dpt. of Energy Economics, Czech Technical University in Prague**
**Prague, Czech Republic**

**Jaromír Vastl**
**Dpt. of Energy Economics, Czech Technical University in Prague**
**Prague, Czech Republic**

## ABSTRACT

The paper describes the results of technical study demanded by the Czech System Operator. The main goal of the study was an improvement of medium term planning process of ancillary services purchasing. The market with these services was quite volatile in the past several years and system operator had major problems with cost planning and budgeting. A substantial volume of the services is bought in advance, usually in three-year period. The important aspect of decision process is whether the current price is a bargain, or whether it is better to wait for more favorable price in the future.

**Keywords:** Ancillary Services, Transmission System Operator, Electricity Pricing, Marginal Cost of Electricity, Regression Analysis.

## APPROACH TO SOLUTION

The task solution was divided in three parts and sub-models:

- average price of services depending upon other economic factors
- shape of supply curve
- marginal costs of services

The average price is derived from regression analysis of the past proposals both bilateral offers and public tenders. The final average price depends upon several micro and macro-economic factors as well as on other mathematic factors of time series. The future price of ancillary services is anticipated with some probability using the forecasted values of previously mentioned factors.

The situation in the market of ancillary services in the Czech Republic is currently not stable. The renewable sources of energy, especially the photovoltaic power plants, create higher demand for ancillary and system services. The higher demand evokes higher marginal and average price of services. The relation between price and volume of service is described by supply curves of their individual services and their shape. This supply curves were modelled on the offers and tenders from the recent history.

Marginal costs of system services are modelled by guessing the cost of power sources supplying these services. The management of electric sources has to opt either for supply of energy or for supply of services. The profitability of service supply has to be at least same as the profitability of energy supply. So the marginal costs of the services can be used as the check of the price of the services. If the marginal price is substantially lower than the offered one, and the price of energy has no market spikes as well, the supplier can be successfully pushed to lower the offered price.

## TECHNICAL AND STATISTICAL ANALYSIS

The authors have received a set of purchase data from system operator (TSO) for following services: primary reserve, secondary reserve, tertiary positive and negative reserve. The technical description of these services is not important for the solution of the task; the services were identically defined over analyzed time period [1].

The basic linear multi-variable regression model for average price modelling has been used [2]:

$$y = a + b_1 \cdot X_1 + b_2 \cdot X_2 + .. + b_k \cdot X_k + e \qquad (1)$$

A simple mathematical calculation (MS Excel) has been used in the first phase of modeling and testing. With regard to the number of necessary models and tests, and with regard to the excessively high demands on the time of model development and analysis, sophisticated statistical software SPSS v. 15 has been taken to use, where designing a model and the necessary analysis and tests was a matter of few seconds.

The objective is to select from all possible factors $X_1$, $X_2$, ..., $X_k$ their appropriate (matching, the best) subset, which is based mainly on variable content examined in service prices. From the methods of subset, selection of explanatory variables can be used as follows:

Forward method – factors in the model are gradually added,
Backward method – factors of the model are gradually withdrawn.
Stepwise method – step regression in which sequence addition or removal of factors are tested.

The choice of independent factors took place with regard to their macro and micro economic value.

PPI Industrial producer price index, the monthly value relative to the month of the previous year [-]. This factor assumption is the inflation dependence, and the possible effects on the price of ancillary services (AnS)

RC Required capacity, respectively required performance for the demanded AnS [MWh]. The price of ancillary services is influenced by the amount of energy demanded (positively or negatively)

ExpPro The index of export production, the monthly value relative to the month of the previous year [-]. The effect of export/import price to AnS price.

Oil Average monthly price of Brent crude oil on the spot market [USD/barrel].

Uranium The average monthly price of Uranium on the spot market [USD/lb]. Energy factors (Oil, Uranium), which may affect the price of service in general.

EEX Average monthly price for an annual supply of electricity on EEX (European Energy Exchange) [EUR/MWh]. Factor, which may affect the price of electricity in general.

Season Seasonal index of given time series related to the average value of the quotation (offered price) [-]. The factor that models the seasonal effects at the rate of the AnS price.

Key Deferring of offered price [month]. Ordinal factor, which objective is to identify the sequence (in time series) of the price.

The number of independent factors and their regression parameters ($b_k$) which represent pricing models of the various AnS, eventuated from the performed analysis and tests of multiple linear regression models.

Analysis of the primary reserve (PR)

ExpPro + Oil + PPI + Season (2)

Analysis of the secondary reserve (SR)

PC + Oil + PPI + Season (3)

Analysis of the tertiary positive reserve (TR+)

Oil + EEX Season (4)

Analysis of the tertiary negative reserve (TR–)

PC + EEX + Uranium + Season (5)

Multiple linear models explain based on the empirically observed parameters of 35.4 % ($R^2$ in the TR+) of the price of required power up to 80.8 % of the price of the tertiary negative reserve. We can say, based on the analysis, that these models sufficiently reflect the settlement prices on the factors considered independent.

Estimate short-term trends in prices of demanded AnS is conducted upon the findings of expected future development of independent factors for 2-3 years. With increasing length of the prediction, uncertainty of prediction increases because of both unexpected macro and micro economic development, as well as with regard to a number of factors used in the pricing model. For these reasons, it is useful to recalculate individual models every two years.

## SUPPLY CURVES FITTING

Many different functions can be used for estimating supply curves. One of the principles, which we have chosen, is an easy interpretation of the curve parameters, and behavior of convex supply curve. Convex behavior means that with the growing demand for services, the price continually increases. Therefore, there is nothing like discounts for the size of service. The existence of a local minimum size is ensured by the convex behavior required for services and vice versa, in high demand, price increases significantly, so it comes automatically to reduce the high demand.

At first calculations, we tried to smooth quotations with several different curves. For some functions was the entire curve, or its significant portion, concave, thus it does not satisfy the above assumptions. At the same time, it proved to be that given the complexity of the task, it would be best to use a function with three parameters, which is simple enough, but it also allows to align offers sufficiently precise.

In the final stage, we have decided to use an equilateral hyperbola to estimate supply curves, more precisely, one of its parts. Three parameters can be easily interpreted, two describe the size of the asymptotes of hyperbola, i.e. a minimum price (for infinitely negative demand) and the maximum possible performance (the exercise price of the service grows to infinity with a value approaching), and the third parameter is the "slope" of hyperbole.

$$Price = \frac{C}{A - power} + D \tag{6}$$

where

A maximum available power of service [MW]
D minimum price of service [CZK/MWh]
C "slope" of supply curve

Supply curves are calculated quarterly, thus every 13 weeks. Each week, a supply curve is determined for each of the services (SR, PR, TR+, TR−) supply curve – offers are sorted by price and offered power volumes are gradually added. It makes no sense to cumulate bids from several weeks together, on the contrary, the resulting curve would not correspond to reality, for it is not possible to move bids between different weeks.

The following procedure was "tuned" in real data so that the calculations were carried out quickly and successfully enough, therefore the method of least squares in a spreadsheet Excel will not have results in a minimum of cases.

Thirteen supply curves for a quarter are smoothed with the hyperbola by the method of least squares. Since this optimization is very complicated, the first guess of the curve parameters of the three points is made. These points are calculated as the average 5th and 15th percentile for power and price, which determines the value estimation of the curve first point. Other points are determined from the 40th and 60th percentile point for the second, and 85th and 95th percentile for the third point. From these three points, an estimation of three parameters of supply curve is made by solving the system of three equations for the three parameters.

The figure 1 presents one example of the final supply curve calculated for the 3rd quarter and the secondary reserve. Blue dots show original data from offers. Red curve shows the final supply curve of the service calculated by the least squares method. Blue and yellow curves represent the maximum a minimum estimates of the supply curve, i. e. the limits for the price of the service.

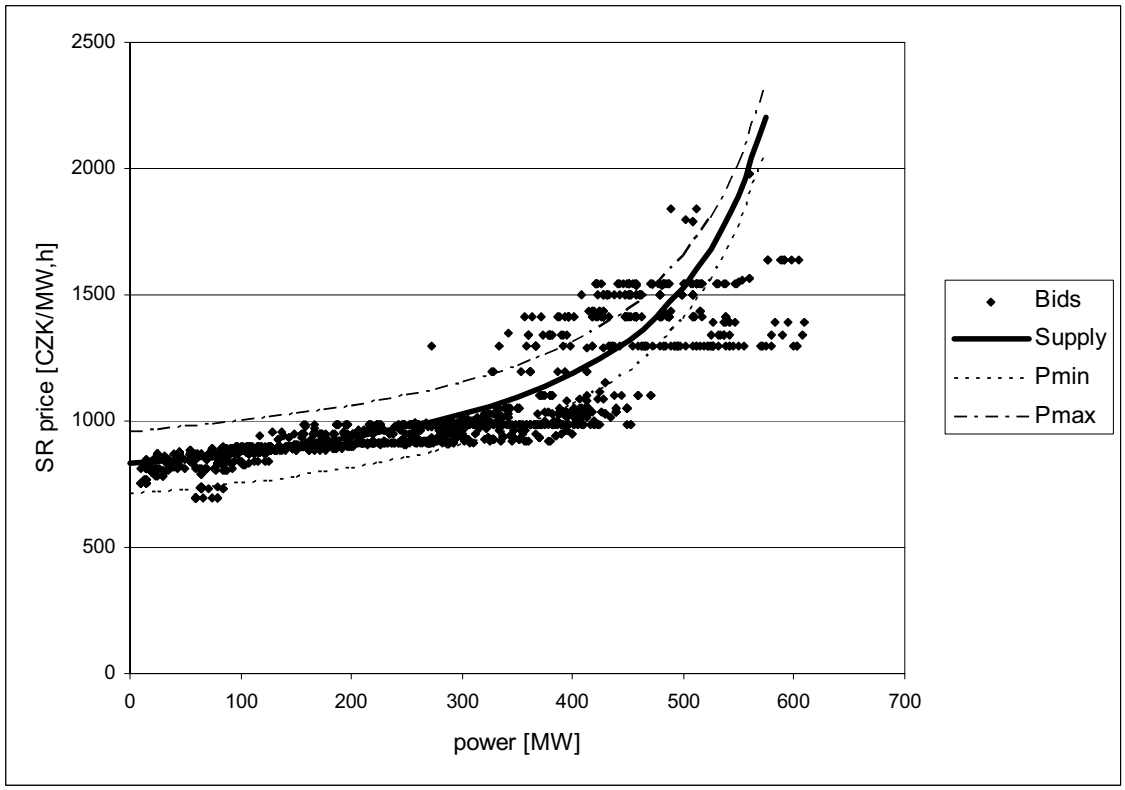



Figure 1. Supply curve for secondary reserve in the 3rd quarter

Curves determined by the procedure in the previous text are not, of course, the final curves, but there are to be adjusted to changing knowledge of the decision maker. The curves parameters will be taken from regression analyze. Curves changed due to changes in economic theory in the neighborhood, described in the chapter "Technical and Statistical Analysis", are calculated from the estimated values of the independent factors.

Another possible change is the shift of the curve according to the knowledge of average price. If the estimate is known, that 300 MW of regulatory power will cost 980 CZK/MWh, the curve can be shifted so that the average price for a given power is met, so that the cost of the service is equal to the required value.

$$D^* = Price^* - \frac{C}{Power^*} \ln \frac{A}{A - Power^*} \quad (7)$$

where

| | |
|---|---|
| A, C | parameters of supply curve |
| $Price^*$ | estimated average price [CZK/MWh] |
| $Power^*$ | estimated purchased power [MW] |
| $D^*$ | new value of D parameter [CZK/MWh] |

The third possible change is the shift of the maximum power, i.e. the vertical asymptote. If a sharp increase in the needs of the regulatory power was expected, an annoying feature of the hyperbole is that after crossing the asymptotes, the functional value falls to minus infinity. For specified purposes only one-half of hyperbole is used, therefore it is necessary to increase the size of maximum offered power (parameter A). At the same time, the other two parameters are recalculated, so that the curve passes through two initial points, relatively determined to the original parameter A (for example, at 0.1 A and 0.5 A).

As the last touch, a curve shift is possible in a situation where a part of the power has already been purchased. The offered value of the power reduces this size and increases the value, because we assume, that the power was purchased at the cheapest possible price.

## MARGINAL COSTS

Long-term marginal costs are the incremental production and transport costs, which are caused by an increase in supply at a given period, which as well includes the investment of producer in the needed expansion of production and transport capacity [3].

Short-term marginal costs are defined as incremental production and transportation costs caused by increase in supply at a given period at unchanged capacity of production and transportation facilities. We can say that it is only a relative increase in operating expenses (excluding depreciation, interest, etc.). Contents of short-term marginal cost is often extended due to a loss caused by possible non-supply of electricity in peak load hours when the supplier may not be getting power to cover the increase of that load.

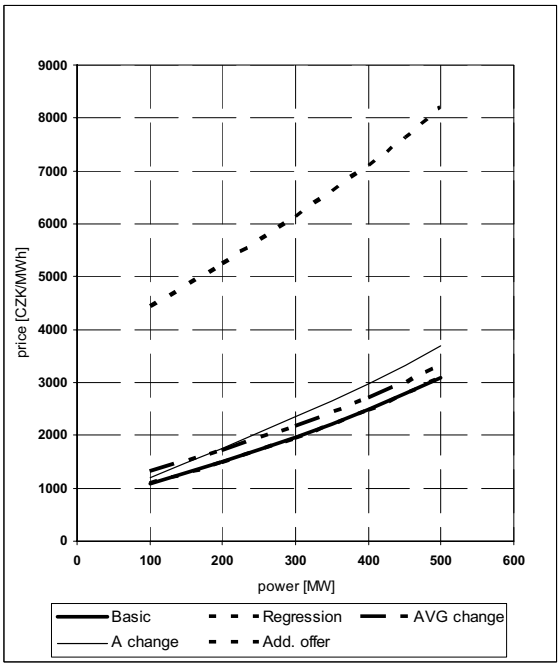



Figure 2. Supply curve with several adoptions

There are two possible approaches to find short-term marginal cost. The first is a calculation based on techno-economic indicators of existing power units [4]. To pinpoint this, it would be necessary to know the parameters of all existing power sources individually, as their fuel and variable costs, as well as their setting in overlaying load diagrams, let us say offered price of its energy market companies. These data are confidential information that cannot be ascertained from publicly available sources.

Another possible approach is to use information from the electricity market and determine opportunity cost. The preferable approach for us is the one described below.

The operator of the power plant can enter a bid for power in the relevant electricity market and obtain unit revenues for sale of electricity of standard base and peak product on the energy exchange, or if needed, by direct bilateral contract. Earning Before Interest, Taxes, Depreciation and Amortization (EBITDA) is obtained by subtracting the average variable costs from these revenues. If the operator considers selling support services, he should get for this service at least the same operating earnings as for sale of electricity. This is the basic idea of this approach with which it is possible to obtain value of support short-term marginal cost services.

Modified approach is under circumstance when the entrepreneur receives payment for the variable component of support services different from its actual variable costs. In a situation where this payment will be less, an increase of short-term marginal cost will occur. If it is on the contrary, there will be its reduction.

Long-term marginal costs represent the upper price limit because they are calculated from the cost parameters of the planned new units and networks, and cover both fixed and variable cost component. On the contrary, short-term marginal cost of support services are the lower price limit because they are derived from the variable costs, and the price level of services is set so that profitability of service delivery is consistent with the profitability of the supply network.

## CONCLUSIONS

The main goal of the technical study was to find the model leading to practical results for system operator and its planning processes. The business data for five years were gathered, checked and analyzed – more than 45 thousand of records. The data were subsequently processed by statistical software, and the main economic factors influencing the prices of the ancillary services were found. The number of these factors was optimized. The final number of parameters is either three or four, the correlation coefficient is from 0.6 to 0.9.

The data were also used to calculate the shape and numerical parameters of the supply curve function. The hyperbolic function was accepted, as this function is convex and suitable for optimization, and the description of its parameters is native. The software module enables to adopt parameters of the function to the changed economic factors, and to the expertise of the users as well.

The marginal cost module is used to check the market prices of the ancillary services to the theoretical value of the services. The long-term marginal costs are the upper limit for the service price because the new sources are used to calculate these costs. The short-term marginal costs are the opportunity costs of sources under current operation, and therefore they are the minimum limit price. The knowledge of the price boundaries involves better purchase timing of services.

The set of program modules was created, and it is used by the Czech TSO. A test for annual operation of the software is expected to be processed, and further cooperation between TSO and authors is planned.

## References

[1] Vastl J., et. al., **Supply Curves of Ancillary Sevices (Cenove krivky podpurnych sluzeb)**. Prague: V-I-P-X (Technical Report), 2010.

[2] Hanke J. E., Reitch A. G., **Understanding Business Statistics.** Richard D. Irwin, ISBN 0-256-06627-2 , 1991.

[3] Beneš M., Starý O., Vastl J., **Marginal Cost of Future Power Units in the CR (Marginalni naklady budoucich vyroben elektriny v CR)** In IVth International Scientific Symposium ELEKTROENERGETIKA 2007 Proceedings. Kosice, 2007, pp. 570-572.

[4] Vašíček J., Knápek J., **Detailed Look on Biomass Utilization Effectiveness - Case Example of the Czech Republic.** In 17th European Biomass Conference & Exhibition. ETA - Florence, Vol. 1, 2009, pp. 2299-2302.

# Fuel Cell Systems for a Greener Aviation

Gwenaelle RENOUARD-VALLET, Josef KALLO, Martin SABALLUS, Florian GORES, Gerrit SCHMITHALS, K. Andreas FRIEDRICH
Institute of Technical Thermodynamics, German Aerospace Center
Hamburg & Stuttgart, Germany

**ABSTRACT**

The trend in new civil aircraft's development is towards "more electric" aircraft and green sky. Fuel cell systems as possible part of the electrical supply in civil aircraft can allow reaching this target. Indeed fuel cell systems have high electrical efficiency, which can be used for on ground power generation and emergency power on flight, and are only producing water and oxygen depleted air as gas emissions. Innovative is that these "waste products" can be used to generate water on board and inert gas for the jet fuel tank as fire retardation and suppression measure. For several years the Institute of Technical Thermodynamics of the German Aerospace Centre (Deutsches Zentrum für Luft- und Raumfahrt; DLR) in cooperation with Airbus developed fuel cell systems which fulfill these functionalities for both aircraft ground and cruise operation. The main achievements of the last past years are the Antares-H2 development and tests as well as the successful milestones passed with the research aircraft DLR ATRA with fuel cell system tests. This paper presents all the modeling and experimental activities of the DLR on fuel cell development for aircraft.

**Keywords**: Fuel Cell, Power, Green, ACARE, emissions, aircraft, system, Antares, ATRA

## 1. INTRODUCTION

A greener sky, that is the vision 2020 issued by ACARE (the Advisory Council for Aeronautics Research) [1]. For several years the Institute of Technical Thermodynamics of the DLR together with Airbus and partners work actively on the development of fuel cell systems for aircraft applications. Fuel cell systems are foreseen as attractive solutions to improve the environmental impact of aircraft systems. Indeed, fuel cell systems can reach 50% of electrical efficiency and advantageous replace power supply systems [2]. The activities of DLR focus on theoretical and experimental investigations regarding specific aircraft relevant operating conditions of PEMFCs (Proton Exchange Membrane Fuel Cell) and HT-PEMs (High Temperature-PEM). In cooperation with partners, several fuel cell applications within the aircraft for both ground and cruise operation could be identified.
In consequence, fuel cell systems capable to support or even replace existing systems have been derived. In this context, the provision of inert gas for the jet fuel (kerosene) tank and electrical cabin power supply including water generation represent the most promising application fields.
This contribution will present the state of development and the evolution discussing the following points:

- Experimental and modeling evaluation of fuel cell systems under relevant conditions (low-pressure, vibrations, reformate operation, etc.).
- Fuel cell tests in DLR's research aircraft ATRA (A320).
- Fuel cell tests in Antares-H2, DLR's flying test bed

## 2. MODELING

Low temperature fuel cell systems in particular PEM-based configurations are currently the only fuel cell type that satisfies the maturity and power density requirements for aircrafts. They can vary significantly in design and functionality. The differences in the individual set-ups arise from specific application purposes as well as from optimization efforts intended to improve the system performance specifically for various aircraft functions. Consequently not all systems would be suited for aircraft applications in the same extend. Based on this background a design study via modeling was carried out to find an auspicious configuration and reduce the number of eligible system setups [3].
The main components of the fuel cell system are the fuel and air supply, the fuel cell stack, the cooling system and the electrical power management (DC/DC, DC/AC converter).
The application oriented results [4] support the positive evaluation of PEMFCs with hydrogen pressure tank based system configurations. A good performance can also be expected for the usage of $H_2$ cryogenic tanks.
The figure 1 gives an overview of the evaluation.

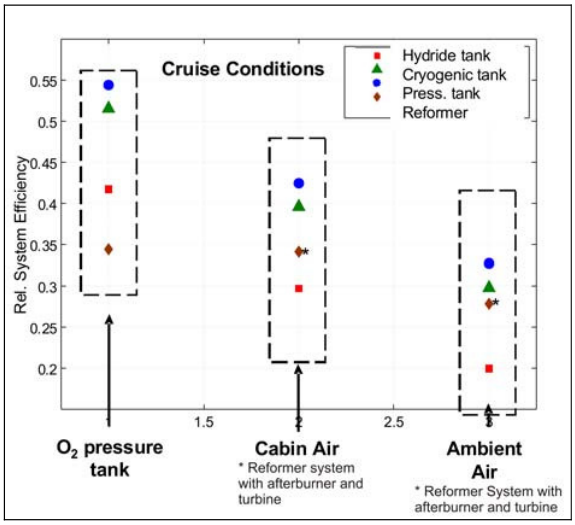



Fig. 1: Relative system efficiencies for considered architectures at cruise conditions [4].

## 3. EXPERIMENTAL EVALUATION IN LABORATORY

DLR has tested cells and systems under aircraft relevant conditions in the laboratory environment. A specific test is the low pressure operation down to 200 mbar which corresponds to a flight height of ca. 12 000 meter (ca. 39000 feet). Even if the final fuel cell system probably will be operated with cabin air, such characterizations are of high importance for system evaluations.

In another test bed, the different possibilities of conditioning the fuel cell cathode exhaust gas are tested in order to optimize the water production for the cabin and the inert gas drying before its distribution in the kerosene tank. Cathode exhaust gas characteristics as mass flow, relative humidity, oxygen concentration, pressure and temperature are evaluated under aircraft representative configurations (pressure and altitude simulated thanks to a vacuum pump) [5].
The water production is ca. 0.5 - 0.6 liter of water per kWh electrical power as shown in figure2.

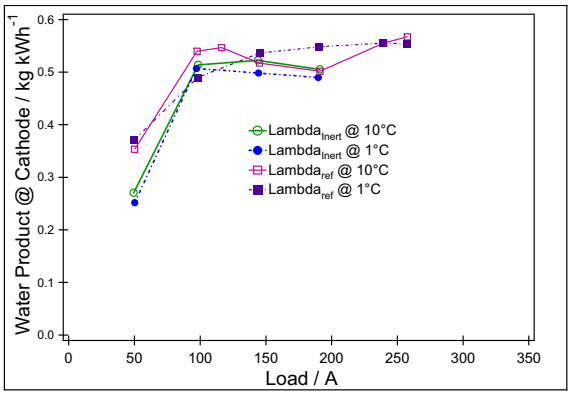



Fig. 2: Water production rate for a commercial PEMFC at different operating conditions and different temperatures of condensation.

By reducing air stoichiometry, it is possible to obtain a low content of oxygen in fuel cell exhaust and to use it as inert gas source. Typical results are presented in figure 3.

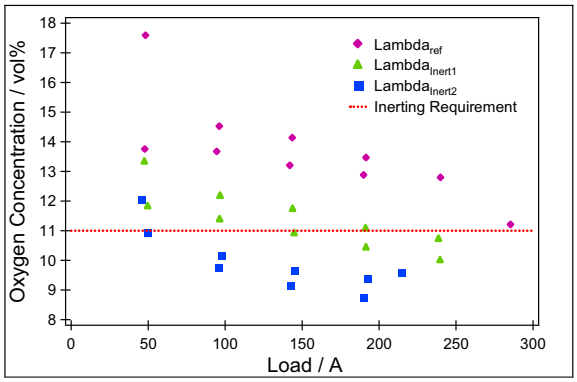



Fig. 3: Inert gas properties for a commercial PEMFC for different operating conditions.

## 4. DEMONSTRATIONS

The development of a multifunctional fuel cell system based on air and hydrogen has to fulfill the rules and regulations of the aircraft industry.
The first functional demonstration was the tests of a qualified $H_2/O_2$ fuel cell system for emergency power in the research aircraft A320 (D-ATRA). The fuel cell system (manufacturer Michelin) was tested in several flights with standardized missions. During the tests, the hydrogen and oxygen based fuel cell system generated up to 20 kW of electrical power. The fuel cell system powered an aircraft electric to hydraulic converter (motor pump) and the back-up hydraulic circuit which also operated the aircraft's ailerons. The system's robustness was confirmed at high gravity loads ("g" loads) during turns and zero gravity aircraft maneuvers. During the flight tests, the fuel cells produced around 10 liters of pure water [6].
Besides emergency power, another system ($H_2$/air) was developed able to deliver inert gas for the jet fuel tank and water during the flight. On ground the system can power on the nose wheel an electrical drive for ground taxiing [7].

In the last 3 years, DLR has developed together with Lange Aviation GmbH company the Antares-H2 – DLR's flying test bed. The Antares DLR-H2 is a high-tech motor glider aircraft based on the Antares 20E self-launching glider, modified by attaching two pods below the wings to carry the fuel cell system and the hydrogen tanks [8-9]. The Antares DLR-H2 flying test bed (figure 4) provides a cost-efficient test environment for developing fuel cell systems, optimizing the test time of the DLR's research aircraft Airbus A320 ATRA.
The advantage of the new test bed consists in the combination of different qualifying test routines. Acceleration, low pressure behavior and climatic environments can be tested in a single run. The fuel cell provides the energy for the electric drive train which consists of electronics, electric motor and propeller. In 2010 an altitude record of 2560 meters was set.

Fig. 4: Antares DLR H2 Flying Test bed

## 5. CONCLUSIONS

Aircraft fuel cell's successful milestone opens the way for future developments like the improvement of energy efficiency and multifunctional operation.
Future research activities will, therefore, focus on control and power management of the fuel cell systems to achieve the required electricity, inert gas and water supply. These goals are partly supported by national projects in LuFo IV (Luftfahrtforschungsprogramm des BMWI, aircraft research program of the German ministry of economics).

On top of these activities the DLR is developing the next generation of fuel cell driven motor glider, Antares H3, in cooperation with Lange Research Aircraft GmbH. The Antares H3 will be a higher-performance successor of the Antares DLR-H2. It is planned to achieve a range of up to 6000 kilometers and endurance of more than 50 hours [10].

## 6. REFERENCES


[1] ACARE report, European Aeronautics: A Vision for 2020, Office for official publications of the European Communities, L-2985 Luxembourg, 2001

[2] S. Eelman, D. L. Daggett, M. Zimmermann, G. Seidel, **High Temperature Fuel Cells as Substitution of the Conventional APU in Commercial Aircraft**, DGLR-JT2003-183, Deutscher Luft- und Raumfahrtkongress 2003.

[3] P. Schumann, C. Graf, K.A. Friedrich, **Modelling and Simulation of a PEM Fuel Cell System for Aircraft Applications,** ECS Transactions, 2008, 12 (1), 651

[4] P. Schumann, C. Graf, J. Kallo, K.A. Friedrich, **Architecture analysis, modelling and simulation of PEM fuel cell system for aircraft application**, Proceedings of H2Expo International Conference and Trade Fair on Hydrogen and Fuel Cell Technologies 2008, Hamburg, 2008.

[5] G. Renouard-Vallet, M. Saballus, P. Schumann, J. Kallo, K. A. Friedrich, H. Mueller- Steinhagen, **Fuel cells for Civil Aircraft Application: On-board Production of Power, Water and Inert Gas**, for consideration for publication in Chemical Engineering Research and Design.

[6] Airbus International Press Release, Emission free power for civil aircraft: Airbus successfully demonstrates fuel cells in flight, 19 February 2008, http://www.airbus.com/en/presscentre/pressreleases/pressreleases_items/

[7] DLR National Press Release: Emission-free airports – DLR develops a fuel cell-powered electric nose wheel for commercial aircraft, 1 February 2011 http://www.dlr.de/en/desktopdefault.aspx/tabid-344/1345_read-28858/

[8] DLR National Press Release: DLR motor glider Antares takes off in Hamburg – powered by a fuel cell, 7 July 2009 http://www.dlr.de/en/desktopdefault.aspx/tabid-5103/8592_read-18278/

[9] DLR National Press Release, Antares DLR-H2 stellt Höhenrekord auf, 21 November 2009 http://www.dlr.de/desktopdefault.aspx/tabid-344/1345_read-21170/

[10] DLR National Press Release Antares H3: DLR and Lange Aviation develop the next generation of fuel-cell powered aircraft, 11 August 2010 http://www.dlr.de/en/desktopdefault.aspx/tabid-344/1345_read-26189/

# Analyzing water supply in future energy systems using the TIMES Integrated Assessment Model (TIAM-FR)

**Stéphanie Bouckaert, Sandrine Selosse, Aurélie Dubreuil, Edi Assoumou, and Nadia Maïzi**
**Center for Applied Mathematics, MINES ParisTech**
**Sophia Antipolis, France**

## ABSTRACT

Even though policies related to water and energy generally dealt with them separately, the two resources are highly interconnected. Energy is required to maintain water supplies while water is essential to produce energy. However, the models developed to estimate future water allocations and energy systems are usually uncorrelated.

In order to jointly consider water and energy uses, a water module has been added to the global TIAM-FR prospective energy system model. Water footprints for the processes in terms of withdrawals and consumptions were implemented in the model. This proposed TIAM-FR model shows how water issues are crucial in energy production and can constitute a tool to observe if future energy mixes could be plausible in regards of water availability.

**Keywords:** water-energy, long-term planning tool, TIAM-FR model

## OVERVIEW

Water and energy are both considered as strategic issues in our societies that have to cope with numerous challenges. Depletion of fossil resources and increasing environmental consciousness are driving the actors of the energy sector towards improving the efficiency of existing technologies and developing new technologies to reduce environmental impacts. Some experts refer to water as the new oil, reflecting a growing concern about the availability and the sustainability of water resources.

While water and energy are highly interconnected, models to estimate the future development of water and energy systems are usually uncorrelated and deal with each resource separately.

- Prospective energy system models may be used to estimate the long-term development of the energy systems and future energy mixes.
- On the other hand, water models may be used to study future allocations of water in sectors such as agriculture, residential, industry, and energy.

To the best of our knowledge, no link between water models and energy system models has been established, which means that the computation of a shared optimum for energy and water resources is unachievable [1, 2, 3]. The current environmental approach in energy issues mainly focuses on reducing greenhouse gas emissions, commonly resulting in increasing use of technologies which provide a partial solution to air pollution issues but may worsen the water stress as they require large quantities of water. In the context of a growing world population with increasing demands and competition for water and energy, it is vital to develop models that include the interactions between the water and energy sectors to compute sustainable long-term management policies.

Water and energy issues have been jointly considered by integrating water allocations for energy purposes into the TIAM-FR prospective energy system model. The proposed model considers the link and dependencies between water and energy, thereby providing plausible developments of the energy system in regions where water is scarce or may become scarce. Water is incorporated in the model in terms of water withdrawal (*"water removed from any sources, either permanently or temporarily"* [4]) and water consumption *("water abstracted which is no longer available for use because it has evaporated, transpired, been incorporated*

*into products and crops, consumed by man or livestock, ejected directly into sea, or otherwise removed from freshwater resources"* [4]).

Water allocations related to the energy chain were implemented in the model according to two methods, a commodity approach and a process approach. The commodity approach relates to every energy process that uses water (mining, refineries, power plants, etc.), while the process approach relates only to electricity production (nuclear, coal power plants… etc.).

## WATER USES

Water is present throughout the energy chain. On the upstream, water is used for extraction, refining and processing. It is also a key factor for the production of non-conventional resources (e.g. oil sands, shale gas, tight gas etc.), currently growing in the oil and gas markets. The quantities of water vary with the method put in place. For example, secondary oil recovery using water injection could require $600m^3$/TJ whereas primary recovery requires only $5m^3$/TJ [5]. The same applies for coal and uranium though quantities of water needed are highly dependent on the type of mine.

| Water uses for power generation | Withdrawal | Consumption |
|---|---|---|
| *Cooling system:* | | |
| - Open loop | + + + | + |
| - Wet closed loop | ++ | + + + |
| - Dry closed loop | 0 | 0 |
| *Heat transfer fluid:* water withdrawals or consumptions related to the main loop of the power plant are negligible compared to these related to the cooling system, therefore this water use will be disregarded. | | |
| *Emission control:* CCS (carbon capture and sequestration) and FGD (flue gas desulfurization) | ++ | ++ |
| *Gasification processes* | + | + |

**Table 1: Water uses for electricity production**

On the downstream of the energy chain, in the electricity production, four major uses of water were identified (see table 1 for an overview). Quantities of water used in the cooling systems are dependent on the kind of technology. As for new processes implemented to reduce the discharge of pollutants or greenhouse gases, they turn out to be high consumers of water: desulfurization systems increase by 10% the volume of water consumed, carbon capture nearly doubles it.

Hydropower has not been considered in this study. Though hydropower plants are by definition the main users of water for energy production, most of these plants do not consume water. Water consumption may be due to additional evaporation from the reservoir (if any), but as dams are also used for multiple purposes (irrigation, flood control, navigation) it makes it difficult to assess which part is related to electricity production.

## THE TIAM-FR MODEL

Analyses carried out in this paper are based on the TIAM-FR developed by MINES ParisTech, Center for Applied Mathematics. The model is based on the TIAM model and the TIMES model framework developed by the ETSAP organization [6, 7]. TIAM-FR is a global model (divided into 15 regions) which offers a technology-rich representation of the energy system. The development of the energy system can be analyzed in a short-, medium-, long-term perspective up till the year 2100. The model is based on a bottom-up, linear-programming approach in which the technical optimum is computed by minimizing the discounted global system cost.

Each step of the energy chain from primary energy to final energy service demands (heating, lighting, travel, etc.) is identified in the model in terms of both economical and technical characteristics. Technologies to achieve these stages are called processes (extraction of fossil fuels, imports, processing of primary energy in final energy etc.). Energy carriers (primary energy, final energy, and useful energy), energy services, materials, cash flows and emissions are called commodities. The links between the commodities and processes are represented in a *Reference Energy System*; an example is shown in the Figure 1 where commodities are represented with an italic font and a light filling, while processes are represented with a normal font and a dark filling.

In the proposed model, water commodities have been linked to extraction, transformation and electricity production (processes framed in red).

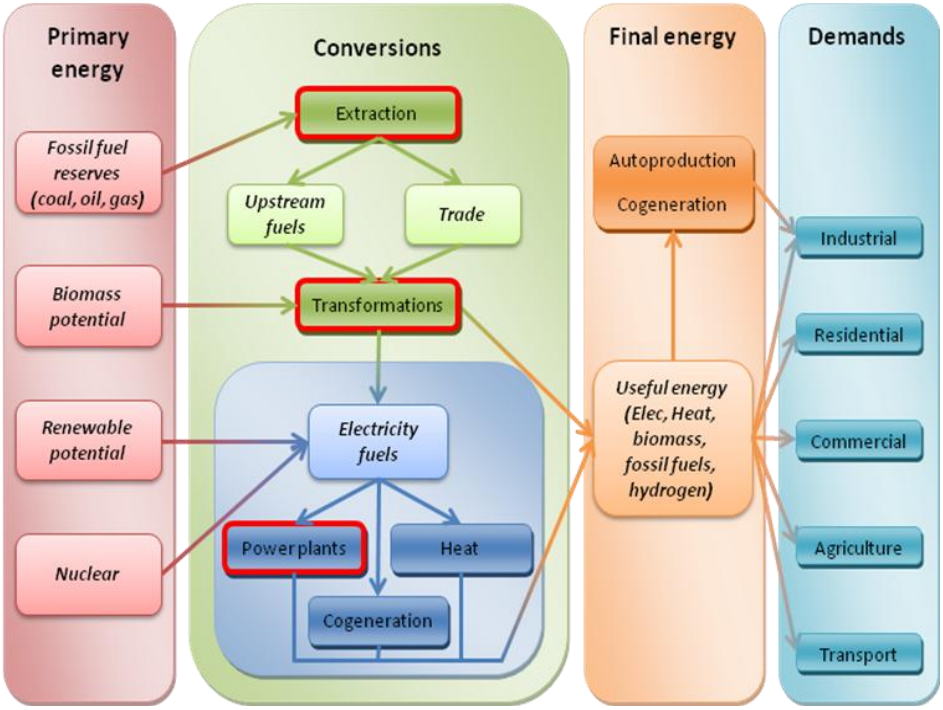



**Figure 1: Reference Energy System**

## COMMODITY APPROACH

The commodity approach may be used to determine water withdrawals and consumptions of the energetic mix. Two types of commodities were linked to each process output, called water factors:

1. Water consumption: *WC
2. Water withdrawal: *WW

The star makes reference to the kind of water used as different kinds of water have been identified for open loop cooling systems while for every other process, only fresh water has been considered:

- B: Brackish water
- F: Fresh water
- G: Ground water
- M: Municipal water
- S Saline water

### Water factors related to the upstream energy chain processes

Data related to water consumption for extraction were found in the literature [5]. Those depend of the kind of technology put in place (type of mines, processes used for oil extraction etc.). However, in the TIAM-FR model, all these processes are not identified so precisely, therefore it has been necessary to weight the water factors according to the different parameters they rely on. Ratios of the various mining methods used for coal and uranium extraction (conventional open pit, underground open pit and in situ leach) were needed to determine water factors for these extracting processes mentioned in the model. Various oil extraction technologies were also identified to adjust their water factors. Other parameters like the percentage of on-shore oil fields or the quantity of *produced water* extracted from oil and gas fields used (thereby reducing the net amount of water injected) were also taken into account.

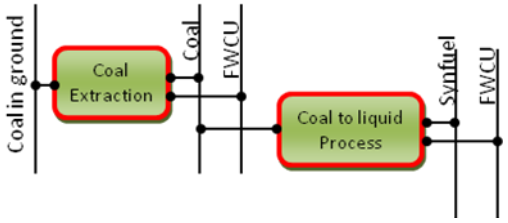



**Figure 2: Upstream water RES with the commodity approach**

Only one output commodity is related to those upstream energy chain processes: FWCU (*Fresh Water Consumption Upstream*), as shown in Figure 2.

**Water factors related to power plants [8]**

The main user of water in a power plant is the cooling system. Withdrawals and consumptions related to these technologies depend on the flow in the cooling system, and thereby on the power plant efficiency. Efficiency varies with the type of power plant (steam turbine, combined cycle etc.) and the fuel used (coal, gas, oil, and biomass). Indeed, the heat to release in the cooling system (Q) is related to the thermal efficiency ($\rho_{th}$) and the mechanical work ($W_m$):

$$Q = \frac{1-\rho_{th}}{\rho_{th}} W_m$$

Water withdrawal ($\omega_{withdrawal}$) needed to release this heat using an open loop cooling system is equivalent to the flow in the cooling system as those do not recycle water:

$$\frac{\omega_{withdrawal}}{W_m} = \frac{1-\rho_{th}}{\rho_{th}} \cdot \frac{1}{c_p \Delta T}$$

where $c_p$ is the water specific heat and $\Delta T$ the authorized increase of temperature. There are no consumptions related directly to the power plant. However, as the EPRI pointed out in one of its studies [9], the increase of temperature is responsible for a substantial evaporation at the downstream of the power plant and may also affect river or marine life. Consumption of open loop systems has been assumed to be 1% of their withdrawal.

Water consumption for wet closed loop cooling systems depends on the rate of heat transfer made by evaporation ($f_{latent}$) and the water latent heat ($h_{fg}$):

$$\frac{\omega_{consumption}}{W_m} = \frac{1-\rho_{th}}{\rho_{th}} \cdot \frac{f_{latent}}{h_{fg}}$$

Withdrawals are needed to maintain a constant flow in the system. Indeed, water losses are caused by evaporation but also by blowdown, as the water evaporated exits the refrigerant in a pure vapor state, leaving minerals behind that will concentrate in the remaining cooling water. The cooling system and the condenser could thereby be damaged. To avoid these issues, a portion of the cooling water has to be removed and replenished with make-up water, which is determined by an acceptable level of concentration (n) [10].

$$\frac{\omega_{withdrawal}}{W_m} = \frac{n}{n-1} \cdot \frac{\omega_{consumption}}{W_m}$$

To allocate a water factor to these power plants, not only their efficiency has to be known but also the ratio of cooling systems (open loop, wet closed loop, dry closed loop) put in place in the different regions available in the model as well as parameters like the allowed increase of temperature.

For some power plants, the implementation of FGD or CCS processes has also to be taken into account, as it will increase substantially their water factors. The same applies to IGCC power plants, as water is used for their gasification process. Some EPRI and DOE[1] studies evaluated water withdrawals and consumptions related to those technologies [1, 9, 11]. Thus, the overall water factor for each process, in each region, takes into account all these parameters and two kinds of water factors, withdrawals and consumptions of different kinds of water were attributed to them, as shown in Figure 3.

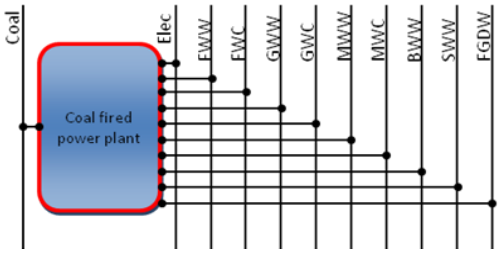



**Figure 3: Electricity production water RES with the commodity approach**

The last commodity, mentioned in the figure, FGDW, represents withdrawals and consumptions (as they were assumed being equivalent) related to the use of flue gas desulfurization.

## PROCESS APPROACH

The process approach applies only to cooling systems for new facilities. Instead of allocating water factors directly to the power plants, a new output commodity, Q, is linked to them, representing the amount of heat that has to be discharged in the cooling system. These new commodities are determined according to the source of heat (combined cycles operating with gas, steam turbines using subcritical coal etc.). New processes related to different cooling systems and different kinds of water are also considered in the model. By requiring that the

[1] US Department of Energy

amount of waste heat (Q) is zero, the model is required to use these processes.

At this level of implementation, the process approach gives identical results as the commodity approach. But further improvements have been made, allowing the model to be more flexible about the concept of water. Indeed, wet and dry closed loop cooling systems induce lower efficiency than open loop cooling systems. Therefore, a quantity of electricity related to the quantity of heat to discharge by the cooling process is needed for them in the model, as shown in Figure 4 .

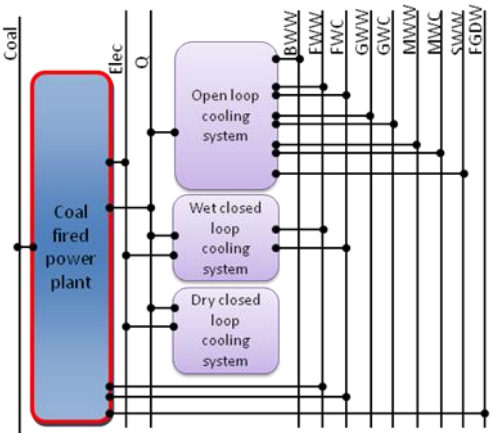



**Figure 4: Water RES with the process approach**

Upper bounds of open loop cooling systems activity have also been established, equivalent to existing processes currently put in place in regard of current incentives for these kinds of cooling systems as well as lower bounds of use of closed loops (wet and dry). It is now possible to leave the choice of cooling systems uses to the model rather than imposing ratios. For a region where water is already scarce, like the Middle East, scenarios of maximum water consumption have been established. With the process approach, the model is able to decide between processes without use of water like photovoltaic or wind energy and the kind of cooling system being used.

## RESULTS

Considering water resources to be unlimited, a *Business as Usual* (BAU) scenario concerning the use of processes was created. The scenario showed the evolution of water allocations; it also showed that worldwide water consumption due to electricity generation may triple by 2050. Additionally, a massive introduction of flue gas desulfurization technologies was found to have a large impact on the water consumption of coal power plants. Such systems might indeed represent 11% of water consumption of these power plants in a scenario where all new coal power plants are installed with these technologies. The impact on water supply with climate constraint scenarios was also studied. These scenarios aim to improve the use of Carbon Capture processes, which increase the power plants water consumption by almost 90%; thus these energy mixes may consume almost four times more water than the current mix. The study further pointed out that even if cooling systems may contribute to high levels of water consumptions, extraction technologies may together represent nearly a third of the water consumption in the overall energy chain.

Stability of the energy mix under limited consumption of fresh water showed that the energy mix is highly dependent on water. An upper limit of water use was defined for each region, according to which the use of processes was evaluated. The constraint on the maximal use of water was found to have a particularly high impact on the electricity generation with the *commodity approach* and the cooling system choice with the *processes approach*. These results consolidate the importance of considering the interaction between the water and energy sectors.

## CONCLUSION AND PERSPECTIVES

Scenarios foreseeing future energy mixes to meet environmental policies commonly do not reflect consequences on the water supply. The study showed that worldwide electricity generation may double by 2050, with an energy mix which may consume almost three times more water than is currently being consumed. As such water consumption levels may prove to be unsustainable; the link between water and energy needs to be taken into account when estimating the development of the energy system.

The proposed TIAM-FR model is a first step towards changing the perception of water in energy models in order to find a global optimum considering both water and energy uses. With this model, it is possible to include environmental considerations and water use throughout the energy chain. Further developments will focus on including other water users (agriculture, residential, industry) in the model in order to perform a global comprehensive analysis of both water and energy supply.

**ACKNOWLEDGMENT**

This research was supported by the Chair Modeling for sustainable development, driven by MINES ParisTech, Ecole des Ponts ParisTech, AgroParisTech, and ParisTech; and supported by ADEME, EDF, RENAULT, SCHNEIDER ELECTRIC and TOTAL.

**REFERENCES**

[1] DOE/NETL. Estimating freshwater needs to meet future thermoelectric generation requirements. 2009 update.

[2] A. Mercè Rio Carrillo and C. Frei. Water : A key resource in energy production. *Energy Policy*, 37:4303–4312, 2009.

[3] *World Water Development Report*. World Bank, 2010.

[4] EEA (European Environment Agency). Multilingual environmental glossary. http://-glossary.eea.europa.eu/EEAGlossary/W.

[5] P.H. Gleick. Water and energy. *Annual review if energy and the environment*, 19:267-299, 1994

[6] R. Loulou and M. Labriet. Etsap-tiam : The times integrated assessment model, part i : Model structure. *CMS*, 5:7–40, 2008.

[7] M. Gargiulo. *Getting Started With TIMES-Veda, Version 2.7*. ETSAP, May 2009.

[8] S. Bouckaert, S. Selosse, A. Dubreuil, E. Assoumou, and N. Maïzi. Implementing water allocation in the tiam-fr energy model. In *11th IAEE European Conference, Energy Economy, Policies and Supply Security: Surviving the Global Economic Crisis, Vilnius, Lithuania*, August 25-28, 2010.

[9] EPRI (Electric Power Research Institute). Water & sustainability (volume 3): US water consumption for power production – the next half century. 2002

[10] SPX Cooling Technologies. Cooling tower fundamentals.

[11] DOE/NETL. Water requirements for existing and emerging thermoelectric plant technologies. August 2008 (April 2009 Revision)

# Industrial and environmental impacts of an expanding bio-energy sector

Nicklas FORSELL
Centre for Applied Mathematics, MINES ParisTech,
Sophia Antipolis Cedex, 06904, France

Ljusk-Ola ERIKSSON
Department of Forest Resource Management, Swedish University of Agricultural Sciences
Umeå, 901 83, Sweden

Edi ASSOUMOU
Centre for Applied Mathematics, MINES ParisTech,
Sophia Antipolis Cedex, 06904, France

## ABSTRACT

In the context of mitigating climate change and increasing energy security, the use of bio-energy production is expected to play a major role. However, an increased use of woody biomass sources for bio-energy production has in Sweden induced pressure on the woody biomass sources available and influenced forestry management. In this paper we analyze the competition of biomass sources and changes in management of woody resources induced by the joint expansion of the bioenergy sector and forest industries. Results show that increased demand of biomass sources may induce a short-term increase of forest harvesting. However, in the long-term, adequate biomass sources was found to be available to fulfill the joint demand of biomass sources for the production of bioenergy and woody products.

**Keywords:** Bioenergy, Biomass, Energy Systems, Forest Industry

## INTRODUCTION

Modern use of biomass for energy purposes commonly spurs from a will to achieve climate goals, decrease greenhouse gas (GHG) emission, support rural development, and increase energy security. As biomass sources can substitute carbon expensive fossil-fuels (e.g coal, oil, gas) for heat, electricity and liquid fuel production, use of biomass is by numerous countries supported as a mean to achieve short-, medium-, and long-term climate targets. Furthermore, as biomass sources may be produced locally, their utilization supports rural development and decrease dependencies of foreign commodity imports, thereby decreasing a countries vulnerability to severe supply disruptions and price shocks.

In many parts of the world, modern use of biomass has expanded rapidly in recent years, and biomass is in more and more countries perceived as a vital source of renewable energy. Particularly in countries with abundant supply of biomass sources, biomass has shown to be cost competitive to conventional fossil-fuels for heating and electricity generation. Most of the biomass currently being used in Europe for energy purposes is forestry biomass, and Sweden is one of Europe's largest users of forestry biomass for energy purposes. In Sweden, biomass and biofuels (including peat, waste, and biofuels such as tall oil, black liquor, ethanol, biogas) supplied 10% (173 PJ) of the total national primary energy use in 1980, and has since steadily increased over the years to in 2009 supply 22% (457 PJ) [1]. Biomass sources are mainly used by the forestry, pulp & paper industries (184 PJ in 2009), district heating sector (180 PJ in 2009), industrial sector (50 PJ in 2009), and by private consumers (41 PJ in 2008) [1, 2]. See Table 1 for an overview of the main biomass sources currently being used by the bioenergy sector.

Table 1: The main biomass sources being used for bioenergy production in Sweden

| Sector | Main biomass sources being used |
|---|---|
| Forest industries | - Industrial by-products (bark, saw-dust, etc.)<br>- Unprocessed forestry residues |
| Pulp & paper industries | - Tall oil<br>- Black liquor |
| District heating | - Wood fuels (logging residues, wood with no industrial uses, recycled wood, short rotation forestry)<br>- Processed forestry residues (pellets, briquettes)<br>- Waste<br>- Peat |
| Industrial sector | - Logging residues wood with no industrial use<br>- Waste<br>- Peat |
| Private users | - Firewood<br>- Processed forestry residues |

As bioenergy production in Sweden is to a high degree depending on forestry biomass sources, the agricultural sector only supplied 5.4 PJ of bioenergy in 2005 [3], the strong growth of the bioenergy sector has induced competition for forestry biomass sources and altered forestry management and harvesting strategies. While the bioenergy sector currently is mostly using forestry biomass sources not in demand by other sectors (e.g. forestry residues, wood with no industrial purposes), it is likely that it will increasingly be relying on biomass sources traditionally used for the production of woody materials and paper. It has for example been observed during the last years that pulpwood has to some extent been sold to district heating plants if they are closer to the harvesting sites than sawmills and pulp & paper factories. Also, harvesting operations have been noted to increase the minimum diameter for pulpwood, resulting in a larger crown and increased production of woody chips for bioenergy production. While the forestry biomass sources available for bioenergy production may diversify in the future (e.g. stump harvesting, intensive forest management), a continued high expansion rate of the bioenergy sector may force the bioenergy sector to rely even further on biomass sources traditionally used for other purposes. Particularly if political instruments are put into place to reach optimistic climate targets, a strong expansion of the bioenergy sector may induce high pressure on the available biomass sources. Sectorial competition for the biomass sources may increase, induce changes in management of biomass source, induce environmental damages, drive up the price of biomass sources, as well as decrease the growth potential of some sectors.

While numerous studies have analyzed the development of the bioenergy sector in terms of the use of food crops for biofuel production [4], or the bioenergy production potential based on freed up and un-utilized agricultural land [5], few have investigated the implications of the growth of the bioenergy sector concerning the use of forestry biomass sources and induced competition of forestry biomass sources [6]. Energy system model that may be used to estimate the development of the bioenergy sector typically only considers the amount of biomass sources available for energy purposes and do not consider the development of other sectors that rely on biomass sources [7, 8]. Related sectors are assumed to develop in an exogenously defined manner even if the supply of biomass sources changes. However, as numerous sectors are dependent on the biomass sources, it is important to jointly consider all sectors in a common prospective methodology, according to which the development of the sectors can be evaluated and the link between the sectors can be analyzed.

The main aim of the paper is to study how the growth of the bioenergy sector may induce competition for forestry biomass sources and changes in forestry management. We particularly focus on how the growth of the bioenergy sector may impact the national forest harvest level, and which forestry and agricultural biomass sources will be used for bioenergy purposes. For this, we use a prospective energy system model that jointly considers the bioenergy sector and forest related industries. The model endogenously computes from the perspective of a least total system cost development of the energy system: which biomass sources will be used by the different sectors, the marginal price of biomass sources, and the resulting national forest harvest level.

## MULTI-SECTORIAL TIMES MODEL

To analyze the impact of an expanding bioenergy sectors on forestry management and forestry related industries, we use a joint TIMES energy system model. The proposed multisectorial model is based on the MARKAL/TIMES modeling framework [9, 10, 11, 12], which are partial-equilibrium, linear-programming models for which the short-, medium-, or long-term development of an energy system can be represented and analyzed. The model is driven by the demand of some specific commodities (e.g. heat, biofuels, electricity, woody products), and is based on a technology rich, bottom-up approach in which typically a large number of technologies can supply the different demands (e.g. 1$^{st}$ and 2$^{nd}$ generation biofuel conversion technologies). The proposed TIMES energy system model jointly considers forestry and agricultural biomass sources, and how the biomass sources may be used to produce woody products (saw material, boards, panels), pulp & paper, district heating, electricity, heat, and biofuels. A range of fully commercial, newly integrated, as well as technologies currently under development are considered in the model to give a full picture of both current and developing conversion technologies.

As the focus of the study is on the forestry biomass sources and the forest industries, the competitiveness of the bioenergy sector were not evaluated within the frame of this report (e.g. electricity production from biomass, nuclear, hydro), nor competition for agricultural biomass sources (e.g. food vs. fuel discussion). While a full range of agricultural biomass sources was considered in the constructed TIMES model (e.g. starch crops, oil crops, grassy crops, woody crops, agricultural residues), only freed-up agricultural land no longer required for food & feed production was assumed to be available for growing energy crops. Improved productivity of food crops as well as technological progress has been forecasted to increase yields agricultural crops and freeing up land for other purposes [13]. Assuming a constant self-reliance level of agricultural crops, only domestic agricultural biomass sources without compromising food & feed production was assumed to be used for growing energy crops.

A full range of forestry biomass sources was considered as available for bioenergy purposes (e.g. forestry residues, pulpwood, saw timber, refined woody products). To accurately consider the flows of commodities between processes, a large number of woody by-products were also considered in the constructed model (e.g. bark, sawdust, wood chips, black liquor, and pulp). See Figure 1 for an overview of the considered flow of commodities for the forest industries. While a large proportion of the woody by-products being produced by forest industries are being used for the production of panels, boards, and paper, significant amounts are also for bioenergy production. Some of the woody by-products are being used directly at the industries to produce heat and electricity for the internal process, however, large quantities are also sold to CHP plants.

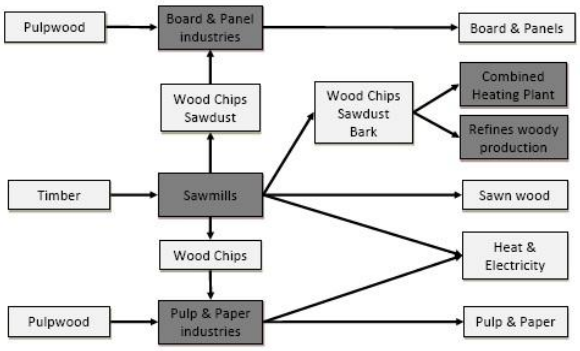



Figure 1: Overview of the considered flow of commodities for the forest industries. White boxes represent commodities while gray boxes represent technologies/processes.

As the demand for biomass sources may affect forestry management, the national forest harvest level was endogenously computed by the model. The harvest level is endogenously computed from a set of exogenously computed applicable and long-term sustainable forest harvest scenarios. Eight harvest scenarios were exogenously created, including a reference harvest scenario (BAU) corresponding to continuation of the current trend in forestry harvesting. The harvest scenarios result in high and low harvest levels during different time periods. As all linear combinations of the exogenously computed harvest scenarios are also applicable and feasible harvest scenarios, the optimal harvest scenario was endogenously computed by the TIMES model as a linear combination of the exogenously defined harvest scenarios. As the rotation age of trees in Sweden is relatively long, 70 years is common, the development of the energy system was studied until 2100 for long-term sustainable forest harvest scenarios to be considered.

To evaluate the sensitivity of the optimal forest harvest level and the development of the Swedish bioenergy system to key parameters, a number of case scenarios were developed. With the developed set of scenarios, sensitivity analysis of model results could be evaluated according to principal economical, technical, and resource assumptions. In all, six scenarios were created concerning the demand of bioenergy and woody products (three concerning demand of bioenergy and three concerning demand of woody products), and five scenarios concerning the cost and availability of agricultural biomass sources (a base case scenario, two concerning cost of agricultural biomass sources, two concerning potential of agricultural biomass sources).

## RESULTS

Overall, the results show that increased competition for biomass sources may affect the national forestry harvest level and induce changes in which biomass source are being used for bioenergy production.

The increased demand of forestry biomass sources was found to be likely to induce a short-term increase in the national forest harvest level (see Figure 2). Increased pressure on woody biomass sources was found to increase forest harvest between the years 2010 and 2035 in comparison to the BAU harvest scenario. After a peak in harvest in 2035, high harvest during earlier periods forced the forest harvest level to decrease below that of the BAU harvest scenario. Thus, in the long-term the computed forest harvest level was found to stabilize to that of the BAU harvest scenario.

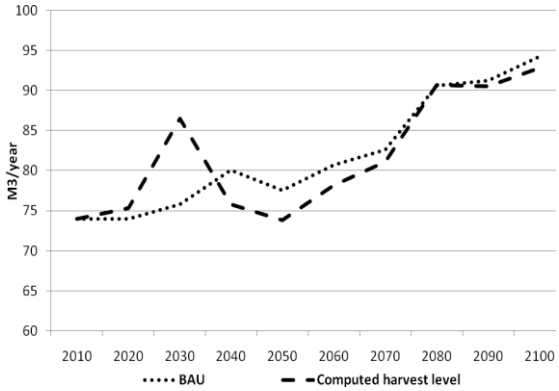



Figure 2: National forest harvest level according to the predefined reference harvest scenario (BAU) and the endogenously computed harvest level computed by the TIMES model (Computed harvest level). Harvest of forestry biomass is expressed in terms of metric tons per year and constitutes the stacked harvest of pulpwood, timber, and forestry residues.

The endogenously computed forest harvest level was found to be surprisingly robust to changes in key parameters. Sensitivity analysis based on the developed case scenarios showed that the computed forest harvest level was stable to the considered changes in demand of bioenergy, supply of agricultural biomass sources, as well as cost of agricultural biomass sources. The computed forest harvest level was only found to be influenced by a decrease in demand of woody products, and this to a fairly moderate degree. If the increment in demand of woody products will be low (approximately 30% lower than in the base case), then the computed harvest level was found to convert to that of the BAU harvest scenario all through the planning horizon.

The induced pressure on forestry biomass sources was also observed to influence the use of forestry and agricultural biomass sources for bioenergy production (see Figure 3). Harvesting of forestry residues was observed to increase from its current level to peak in the year 2025. However, after 2025 the use of forestry residues for bioenergy purposes was observed to decrease, and the increasing demand of biomass sources for bioenergy production was instead supported by woody crops grown on freed-up agricultural land (e.g. willow, poplar). Large amounts of woody crops were estimated to be grown on agricultural land after 2020 to support an increased demand of biomass sources for the production of woody products and bioenergy. Thus, after 2020, the bioenergy sector starts using a large proportion of woody crops instead of forestry residues as feedstocks for electricity and heat production. Note however, that the decrease in use of forestry residues for bioenergy purposes was to a high degree due to a decrease in the availability of forestry residues and not by a decrease in the share of available forestry residues being harvested. Forestry residues were assumed to only be collected from final felling, and as the national harvest level decreases, so does the availability of forestry residues.

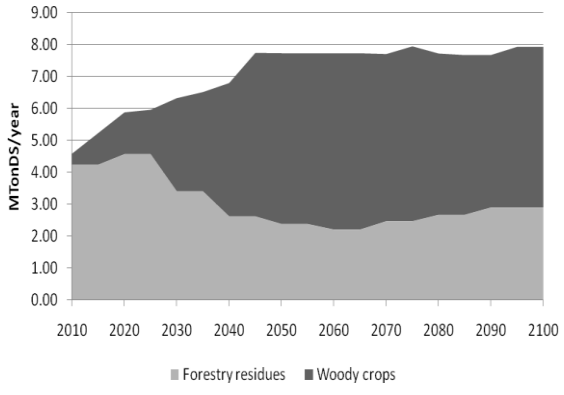



Figure 3: Harvest of woody crops and forestry residues for bio-energy purposes (heat & electricity). Harvest of biomass sources is expressed in terms of Million of tons of dry biomass per year.

## DISCUSSION

While the increased joint demand of biomass sources by forest industries and the bioenergy sector was found to be likely to induce a short-term increase in the national forest harvest level, it did not induce a long-term deviation from the current trend of forest harvesting. The increase in forest harvesting was still relatively small and only lasted over a short period of time. As the amount of freed up agricultural land from food and feed production increases, a large amount of energy crops may be grown on agricultural land, increasing the supply of biomass sources. In the long-term, the harvest level stabilized to that of the reference harvest scenario corresponding to a continuation of the current trend in forest harvesting. This indicates that the impact on forestry management may be significant in the short-term, but less significant in the long-term. An adequate and reliable supply of biomass sources can be materialized to support both bioenergy generation and the needs of forest industries.

However, the results show that a short-term increase is the forest harvest level would be beneficial for the joint development of the bioenergy sector and forest industries. If supply of biomass sources is not mobilized fast enough to match demand, this may lead to shortages in pulpwood, and price of biomass sources may rise even as production costs continue to fall. However, as logging and transportation techniques for forestry biomass are well developed, and as the logistic chain from harvesting sites to factories is well mature, Sweden is in a good position for performing such an increase in forest harvesting. Forestry biomass sources could thus act as a supply buffer until the production of energy crops has been increased.

It is important to consider that the potential forest harvest level is only a theoretical measure, and a realization of the harvest level is dependent on a number of factors. In actual conditions, the timing of forest harvesting operations is decided by forest owners, whom may require further incentives to increase harvest rates. As the current national forest harvest level is much lower than the annual forest increment, economical incentives may have to be put into place to increase harvest levels. Such incentives would help to safeguard the supply of biomass sources at a price that support the development of the bioenergy sectors and forest industries.

Furthermore, steps may have to be taken to ensure the production development of energy crops. Production in Sweden of energy crops has been modest so far and numerous obstacles need to be overcome for a large scale production to mobilize. Significant developments are required in terms of supply infrastructure, production chain, and distribution chain. There is also a considerable amount of uncertainty around the attractiveness of energy crops,

an uncertainty that creates hesitation towards energy crops on the parts of companies and farmers. Farmers may not be willing to switch from annual crops to growing energy crops with a rotation period of 20 to 30 years until instruments are put into place to guarantee sufficient demand at a profitability level 10 to 15 years from now. Until a stable and sustainable demand of energy crops is proven, farmers may be unwilling to switch to energy crops that take two harvest cycles to recoup initial investments and that only generate income every three to five years.

While heat and power generation from biomass sources can reduce GHG emission, production and logistics need to be performed in a sustainable manner. Large-scale biomass production of energy crops will require new legislations and the setting of industrial standards for biodiversity and other environmental values not to suffer. It is difficult to predict the potential impact on environmental and biodiversity values by a short-term increase in forest harvesting. Increased harvest operations may inflict damage to key animal habitats, and as the age structure of the forest will be affected by the increased harvest, the amount of available habitats of some species may be affected. The supply of dead tree biomass in the forest may affect the forest ecosystem. However, biodiversity issues are commonly considered when logging operations take place. Legislations and industrial standards to prevent loss of biodiversity have been developed over a long time and may be considered as mature. Also, as Swedish legislation does not allow deforestation, the impact on environmental and biodiversity values may be minor. Furthermore, as a decrease in forest harvest would take place after the peak year, the resulting forest volume in 2100 would be the same as if forestry harvesting would continue according to the current trend.

## CONCLUSIONS

This study indicates that the demand for forestry biomass sources for the production bioenergy and woody products may induce a short-term increase in the Swedish forest harvest level. This calls for better co-ordination of forestry, agricultural and energy policies to make sure that industrial and environmental values are not hampered. Although biomass production can be increased significantly in the near future both in terms of increased land and feedstocks available for bioenergy production, it is nonetheless a limited resource.

## ACKNOWLEDGEMENTS

This research was supported by the Chair Modeling for sustainable development, driven by MINES ParisTech, Ecole des Ponts ParisTech, AgroParisTech, and ParisTech; and supported by ADEME, EDF, RENAULT, SCHNEIDER ELECTRIC and TOTAL.

## REFERENCES

[1] "Energy in Sweden - facts and figures 2010", **SEA (Swedish Energy Agency)**, No. ET 2010:45. 2010.

[2] "Swedish Statistical Yearbook of Forestry". **Swedish Forest Agency**, 2010.

[3] "Bioenergi från jordbruket - en växande resurs (Bioenergy from agriculture - a growing resource)", **Statens offentliga utredningar (Official Government Report)**. SOU 2007:36, 2007.

[4] R. T. Govinda, A. Shrestha. "How much hope should we have for biofuels?" **Energy**, 2010, pp. 1-15.

[5] G. Bernades, M. Hoogwijk, R. Broek. "The contribution of biomass in the future global energy supply: a review of 17 studies" **Biomass and Bioenergy**, Vol. 25, No. 1, 2003, pp. 1-28.

[6] M. Bisaillon, P. Fritz, B. Rydén, P. Erik Springfeldt, T. Unger. "Biokonk: Konsekvenser för energi- och skogssektorn av förändrad efterfrågan på biomassa" **Elforsk rapport**, No. 08:61, 2008.

[7] "EU27 Synthetis report deliverable". D.4.2. **Intelligence Energy Europe**; 2009 Aug. Report No.: EIE/06/170/SI2.442662.

[8] M. Börjesson, E.O. Ahlgren. "Biomass gasification in cost-optimized district heating systems–A regional modelling analysis", **Energy Policy**, Vol. 38, No. 1, 2010, pp. 168 – 180.

[9] L. Fishbone, H. Abilock, "MARKAL, a linear-programming model for energy systems analysis: technical description of the BNL version" **International Journal of Energy Research**, Vol. 5, No. 4, 1981, pp. 353–375.

[10] A. Kanudia, R. Loulou. "Advanced bottom-up modelling for national and regional energy planning in response to climate change", **International Journal of Environment and Pollution**, Vol. 12, No. 2, 1999, pp. 191–216.

[11] A. Kanudia, M. Labriet, R. Loulou, K. Vaillancourt, J.P. Waaub. "The World-Markal Model and Its Application to Cost-Effectiveness, Permit Sharing, and Cost-Benefit Analyses", In: **Energy and Environment**. R. Loulou, J.P. Waaub, G. Zaccour, editors. Springer; 2005. p. 243-268.

[12] M Gargiulo. "Getting started with TIMES-VEDA. Version 2.7." **Energy Technology Systems Analysis Program (ETSAP)**. 2009.

[13] M. D. A. Rounsevell, J. E. Annetts, E. Audsley, T. Mayr, I. Reginster. "Modelling the spatial distribution of agricultural land use at the regional scale", **Agriculture, Ecosystems & Environment**, Vol. 95, No. 2-3, 2003, pp. 465 – 479.

# INDUSTRIAL HEAT RECOVERY AND HEAT PUMPS SYSTEMS

***Gondia Sokhna SECK***,
Center for Applied Mathematics
Ecole des Mines de Paris (MinesParisTech)
BP 207, 06904 Sophia Antipolis cedex, France
< gondia.seck@mines-paristech.fr >

**And**

***Gilles GUERASSIMOFF***
Center for Applied Mathematics,
Ecole des Mines de Paris (MinesParisTech),
BP 207, 06904 Sophia Antipolis cedex, France.

**ABSTRACT**

The growing costs of fuels and supply security concerns make it necessary to reduce energy consumption in industrial processes. The use of heat recovery technologies like Heat Pumps (HPs) is an effective way of achieving an energy saving and a significantly reduction in $CO_2$ emissions. Around one third of the final energy used for thermal purposes[1] is subsequently wasted through losses. The lowest temperature range of this heat could be technically recovered through HPs systems from energy end-uses like dryers, refrigeration systems, air compressors etc. HPs have made great technological progresses by providing at the same time useful heat at higher temperatures and the possibility of replacing boilers which consumed traditional fossil combustible to provide heat. The purpose of our work is a prospective energy analysis up to 2020, to highlight the availability and opportunities of existing and innovative high temperature heat pumps systems in French Food & Drink (F&D) industry. For that purpose, we implement a sectoral energy system optimization model using the ETSAP TIMES[2] framework. It is a "Bottom up" technical economic model which provides a technology rich basis for estimating energy dynamics over a medium or long-term and a multi-period time horizon.

**Keywords**

Energy policy, Climate change, Heat Pump system, Heat recovery, Industrial waste heat, energy efficiency, TIMES.

## 1. INTRODUCTION

Almost three-quarters of final energy consumption in industries are used for thermal purposes (boilers, thermal end-uses) and this heat is mainly produced by combustion of fossil fuels which generate great amounts of $CO_2$ emissions. However, around one third of the final energy used for thermal purposes is subsequently wasted through losses (fumes and waste water).

Under Energy Savings Certificate (ESC or White certificates), some measures such as boilers economizers and variable speed drives have been applied because of their high potential of energy efficiency and simple implementation on industrial site. However, estimating precisely energy savings for more complex projects is not simple. Between 10% and 30% of final energy consumption are in theory possible to recover in the losses on thermal equipments which represent around 30 to 100 TWh per year for France in industry. Thus, our modeling problematic is heat recovery on industrial processes which could be developed and recognized as eligible for ESC.

The lowest temperature range of this heat could be technically recovered through heat pumps (HPs) systems from thermal energy end-uses like dryers, ovens…etc. But we can recover refrigeration systems, air compressors etc... They can be used to raise the temperature of waste heat so that it can be also re-used for processes and space heating purposes. However, it is seldom re-used today. HPs have made great technological progresses by providing at the same time useful heat at higher temperatures and the possibility of replacing boilers which consumed traditional fossil combustible to provide heat. Therefore, it seems important to recover and re-use this heat in order to achieve the 20/20/20 EU scheme against climate change for an energy efficient and low-carbon economy. Heat pumps represent an important technology which can significantly reduce $CO_2$ emissions and provide attractive opportunities for energy conservation in industry. But their adoption will depend mainly on their economical competitiveness.

The purpose of our work is a prospective energy analysis up to 2020, to highlight the availability and opportunities of existing and innovative high temperature heat pumps systems in French F&D industry which considered as Non-Energy Intensive (NEI) industry [1]. We focused on this sector by his energy and economical importance in industry. This sector represents around 13% of total industrial value added for around 15% of total final energy consumption in French industry with about 63 TWh for 2005 [2][3]. It is the third greatest energy consumer sector after iron & steel sector and chemistry sector.

Heat is very important in food processing (drying, baking, pasteurization, heat sterilization, blanching...) which will allow destroying enzymatic or microbiological activity or removing

[1] Almost three-quarters of final energy consumption in industries are used for thermal purposes (boilers, thermal end-uses) and this heat is mainly produced by combustion of fossil fuels which generate great amounts of $CO_2$ emissions.

[2] The Integrated Markal-Efom System.

water to inhibit deterioration. We observed that 85% of the total heat demand is consumed by only four energy end-uses (dryers, heating liquid and gazes, heat treatment and evaporation concentration). And around 95% of this heat needed which means approximately 25 TWh is between 60°C-200°C. The heat demand on the 60-100°C range represents almost 45% while in the 100-140°C; the heat demand is around 40% of the total. Three subsectors such as Dairy, Starches and Sugar consumed 55% of the total heat demand of French food and drink[3].

## 2. THE TIMES MODEL

We implement a sectoral energy system optimization model using the ETSAP TIMES framework. TIMES is developed and maintained by the Energy Technology Systems Analysis Programme (ETSAP), an implementing agreement under the aegis of the International Energy Agency (IEA). TIMES is an economic linear programming model generator for local, national or multi-regional energy systems. It is a "Bottom up" technical economic model which provides a technology rich basis for estimating energy dynamics over a medium or long-term and a multi-period time horizon. It is usually applied to the analysis of an entire energy sector like industry, but it may also be applied to study in detail single sectors, like the Food and Drink industrial sector here (Fig. 1) [4].

*Source Uwe Remme*

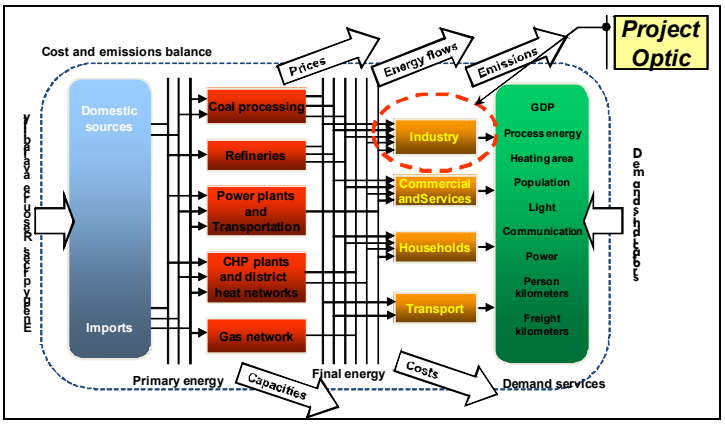



**Fig. 1 :** Overview of TIMES model.

The TIMES objective is to minimize the total cost of the system which includes capital costs, Operation & maintenance costs, exogenous imports or exports, taxes and subsidies…etc. All costs elements are discounted to a chosen reference year. Furthermore, TIMES is a partial equilibrium computation on energy markets. This means that the model computes both the flows of energy forms and materials as well as their prices, in such a way that, at the prices considered by the model, the suppliers of energy produce exactly the amounts that the consumers are willing to buy [5]. The cost discounted explains why we have to choose a global discount rate (see after) but we can specify it for each technology such as heat pump

[3] The Food & Drink industry is subdivided in 20 subsectors according to the Statistical classification of economic activities in the European Community (NACE).

It is based on a Reference Energy System (RES) which is a network describing the flow of commodities through various and numerous processes [6]. The energy description has to be done by process step because the process flows could be fairly well defined for a single broad product line by unit process step (iron and steel, paper and allied products, glass and glass products…). This technological description is well suited to Energy Intensive (EI) Industries. However, in the case of NEI Industry like F&D sector, this method is difficult to apply, and so it requires another approach because of the diversity and the large number of end products and unit processes. This explains our choice to develop a modeling approach by energy end-uses (e.g. drying, heat treatment…) (Fig. 2). In this work, we assume eleven energy end-uses which represent an aggregation of existing unit processes (we can see some examples in Fig. 2) in food and drink sub-sectors.

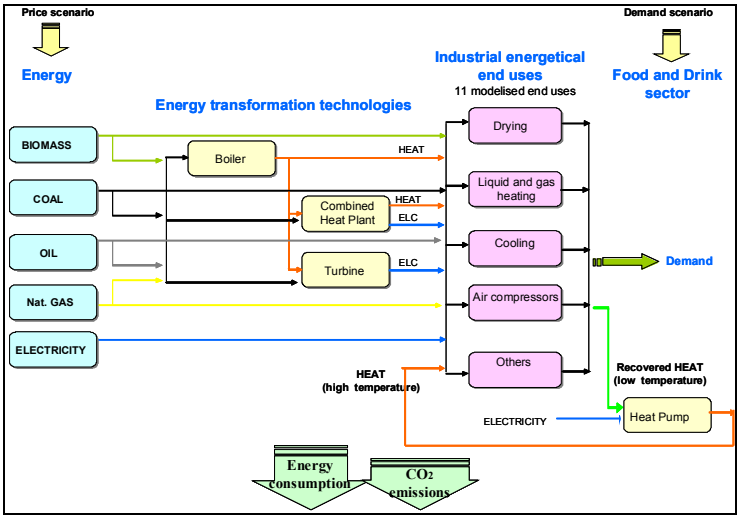



**Fig. 2 :** Reference Energy System for Food and Drink sector.

This approach allows us to build a generic model unlike the energy intensive industry where we have to model each sector differently. By adding the amounts of energy consumed in each energy end-use, we can calculate the total energy consumption of each industrial sub-sector. We have to define four types of input to obtain a complete scenario in TIMES. These inputs represent energy service demands, primary resource potentials, a policy setting, and the descriptions of all technologies (efficiency, all costs such as investment cost, O&M cost…etc)..

## 3. DESCRIPTION OF DIFFERENT HYPOTHESIS AND SCENARII

We considered a highly disaggregated level (4-digit level of NACE classification) of the Food and Drink industry, such as 35 sub-sectors.

- ***Heat pump systems***

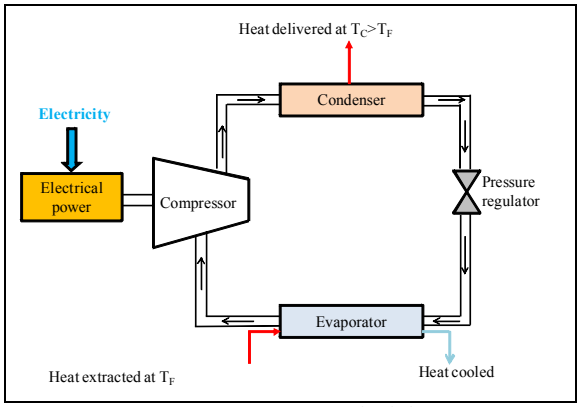



**Fig. 3 :** Heat pump principle.

From the first law of thermodynamics, we can define the relation between the amount of heat extracted $Q_F$ at the temperature $T_F$, the amount of heat delivered $Q_C$ at the temperature $T_C > T_F$, and the mechanical energy W supplies to the system:

$$(1) \quad W + Q_F = Q_C$$

The theoretical Coefficient of Performance (the Carnot COP) which is the efficiency of the HP is defined by:

$$(2) \quad COP_{theoretical} = \frac{Q_C}{W} = \frac{Q_C}{Q_C - Q_F}$$

We can deduce from the Clausius equation, the relation between the theoretical COP and the different temperature for an ideal thermodynamics HP:

$$(3) \quad COP_{theoretical} = \frac{T_C}{T_C - T_F}$$

But the COP of real systems which are irreversible is around the half of the values obtained in the (3) relation [7]. And so, we assume that the real COP is defined from the Carnot COP with a factor according to industrial experts [8]:

$$COP_{real} = 0.55 * COP_{theoretical} = 0.55 * \frac{T_C}{T_C - T_F}$$

In this paper, we will subdivided in seven temperature ranges between 60°C-200°C the heat demand in the F&D subsectors, such as four 10°C ranges between 60-100°C, two 20°C ranges between 100-140°C and the last range 140-200°C. Heat pumps are currently providing heat up to 100°C. We have many developments for high temperature heat pump in many laboratories. We have the example of EDF which works with several French R&D programs for the development of industrial high temperature heat pumps providing heat up to 140°C [7]. First prototypes will be tested in Lab perhaps in these two next years. We considered this generation of HP will be on market in three or four years from now. This analysis allowed seeing the possibility of their penetrations. Their investment costs are 20 or 30% more expensive than those of the heat pump up to 100°C.

According to the temperature of heat extracted at 45°C as we said before, we can calculate the different COP for different HP in each temperature range:

| **Temperature ranges** | **60 to 69° C** | **70 to 79° C** | **80 to 89° C** | **90 to 99° C** | **100 to 119° C** | **120 to 139° C** |
|---|---|---|---|---|---|---|
| **COP** | 6,29 | 4,85 | 3,99 | 3,42 | 2,85 | 2,36 |

**Table 1 :** Considered COP for Heat pumps

Their adoption will mainly depend on their economical competitiveness with their investment costs and the evolution of energy prices up to 2020.

➢ ***Heat recovery opportunities on energy end-uses***

In F&D industry, heat is mostly wasted with the temperature between 30 and 60°C in the industrial energy end-uses. So in this modeling, we considered an average temperature of 45°C. This heat temperature is too low to be recovered by an exchanger and re-used in industrial end-uses. So, HPs can be used to raise the temperature of waste heat so that it can be also re-used for processes and space heating purposes. We distinguished, as we said before, eleven energy end-uses.

Heat recovery on air compressors

Compressed air represents around 7% of total electricity consumption in French F&D. It is a major energy end-use in industry. An estimation of over 90% of the input energy (electricity) to air compressor is lost as waste heat [9].

*Source GPG 238*

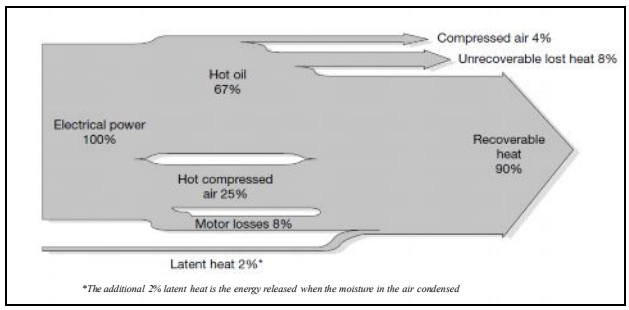



**Fig. 4 :** Energy flow in air compressors (example of oil injected screw compressor).

But if we assume that almost 5 to 20% of additional losses is possible in the case we have an extensive or short ductwork to the HP, this estimation goes down between 70% and 85% of the input energy which can be recovered as heat. In this paper, we used a heat recoverable in air compressor of 70% in pessimist view.

$$Heat\ recoverable_{Air\ compressors} = 70\% * Energy\ input$$

Heat recovery on chiller condensers

Cooling is also a major end use in industry and represent almost one quarter of total electricity consumption in F&D industry. The heat available at their condenser is generally wasted and the calculation of heat recoverable is based on the first law of thermodynamics and the definition of the energy efficiency ratio (EER). Thanks to the mechanical energy W supplied to this system, E the electric input power, we absorb to the heat source in the temperature thermodynamics $T_F$ with the heat

energy $Q_F$ and we reject the heat at the condenser to the temperature $T_C$ with the heat energy $Q_C$:

$$W + Q_F = Q_C \quad \text{First law of thermodynamics}$$

$$EER = \frac{Q_F}{E} \quad \text{Definition of the EER of the machine}$$

We deduced from these two equations[4] the total heat which is lost at the condenser:

$$Q_C = (1 + EER) * E$$

In this paper, we assume, according to industrial experts at EDF R&D, the chiller EER during an entire year is estimated around 2.5 and we also considered that only 70 % of $Q_C$ is recoverable by heat pumps [10].

$$Heat\ recoverable_{Chiller\ condenser} = 70\% * (1 + 2.5) * Energy\ input$$

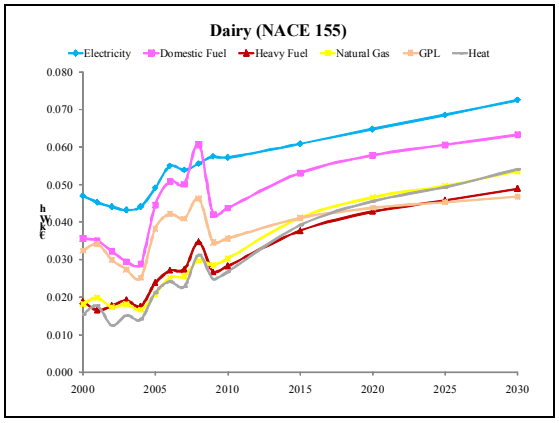
Heat recovery on other thermal end-uses

Due to the difficulty to obtain data at the end-uses level, we assume that for all other end-uses unless air compressors or end-use of cooling and refrigeration, the heat recoverable is between 15%-25% of total input thermal end-uses in average with industrial experts at EDF R&D. In some studies, we observed that this waste heat could be more important up to 55% like in some US industries [11]. Indeed, we assume a factor which is 15% meanwhile having an accurate estimation for each end-use due to the good French policy in energy efficiency in industry.

$$Heat\ recoverable_{Other\ thermal\ end-uses} = 15\% * Energy\ input$$

- ***Energy prices scenario***

We built our scenario by starting from the real energy prices noticed historically (between 1993 and 2009) in each subsectors of NEI industry.



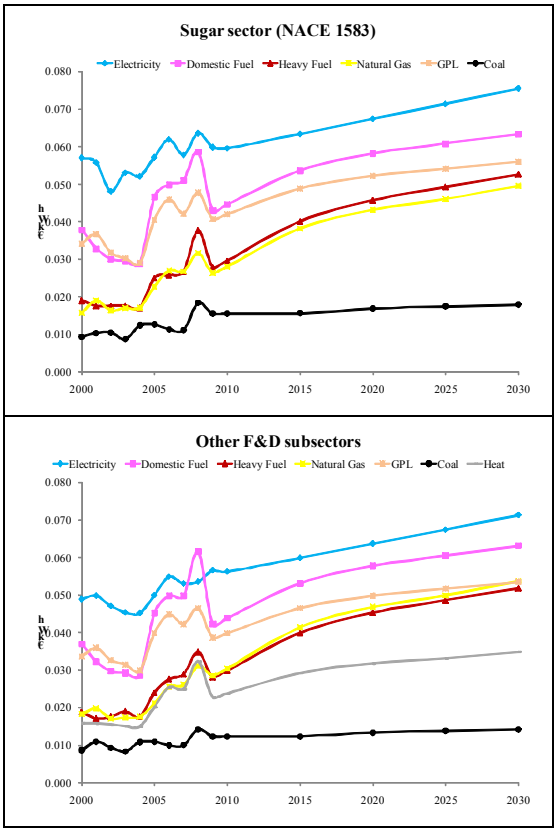




**Fig. 5 :** Scenarii of energy prices for all subsectors in French F&D industry.

Indeed, the projections of energy prices stemmed from models such as POLES[5] or the World Energy Outlook 2010 reflect more those of the EI industries. The energy contracts signed by NEI industry with energy operators are totally different to those of EI industry due to the energy consumption weight. The hypothesis which we made is to start at the real levels of historic energy prices used in the NEI sectors, then to prolong them by using the same projections as IEA with specific elasticity.

## 4. RESULTS

To know the impact of HP on the final energy consumption and the CO2 emissions in French F&D industry, we compare two scenarios.

The first scenario represents the Business-As-Usual Scenario (Sc_BAU), in which no change in policies is assumed. This scenario could be called the Reference Scenario, is attended to serve as a baseline against which the impact of new policies can be assessed. The second scenario is the Heat Pump Scenario (Sc_HP), in which we assumed that heat pumps are on the market and could be used for heat recovery for an energy efficiency policy.

- ***Evolution of final energy consumption***

[4] For simplification, we assume that it is a perfect compressor with no loss. And thus the electric input power E is equal to the power transferred as work to the refrigerant.

[5] Prospective Outlook on Long-term Energy Systems, developed by the LEPII (A research lab in economy and energy policy) which is at Grenoble (France)

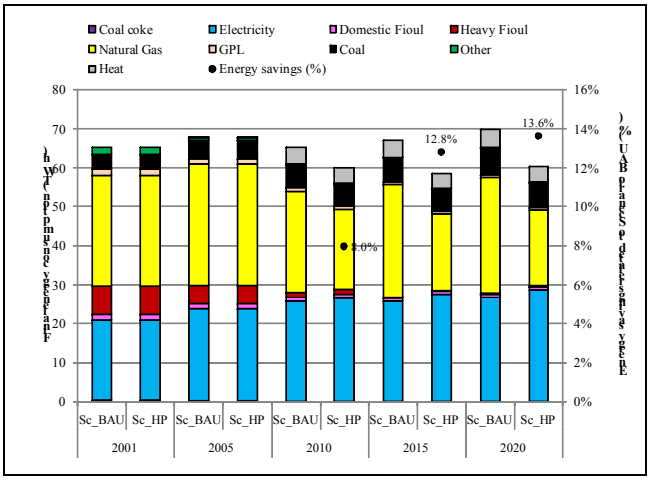



**Fig. 6 :** Impact of HP on the evolution of final energy consumption in French F&D up to 2020.

Finally, we obtained an evolution of the final energy consumption about -7.5 % relative to the level of 2001 (-12.1 % relative to the level of 1990) thanks to the deployment of HP in F&D industry, i.e. approximately 60.35 TWh up to 2020. On the other hand, without the implementation of these HP in F&D industry, we would have obtained an increase of the final energy mix of 7.1 % compared with 2001 (+1,7 % compared with 1990) to reach 69,9 TWh.

**We finally achieve around 9.5 TWh of energy savings up to 2020 which represent approximately 13.6 % of the final energy consumption in French F&D industry** (Fig. 6). It is distinguished by subsector for a better screening of the impact of HP in F&D industry (Fig. 7). These energy savings corresponds to a substitution of around 2 TWh of electricity against approximately 11,5 TWh of fuels (90,8 % of natural gas, 3,7 % of coal and 5,4 % of bought heat).

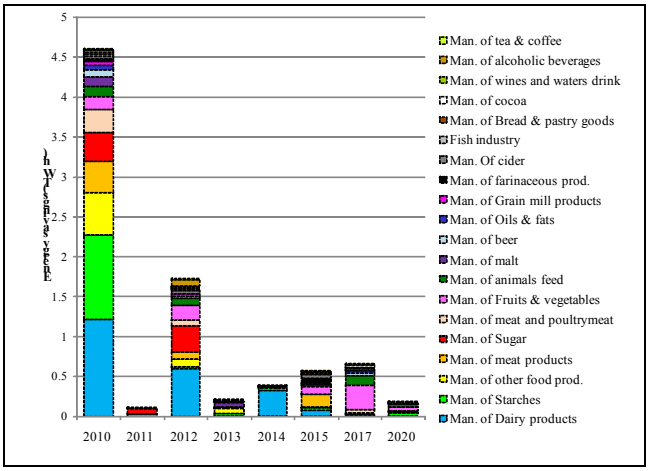



**Fig. 7 :** Evolution of energy savings by subsector in french F&D industry.

- ***Range temperature economically achieve by subsectors***

Fig. 20 below present the levels of heat demand, the heat production of HP and their penetration rate as well as the temperature ranges reached at a subsectoral disaggregation in French F&D industry up to 2020. We grouped all subsectors in three graphs presented from the highest heat demandto the lowest.

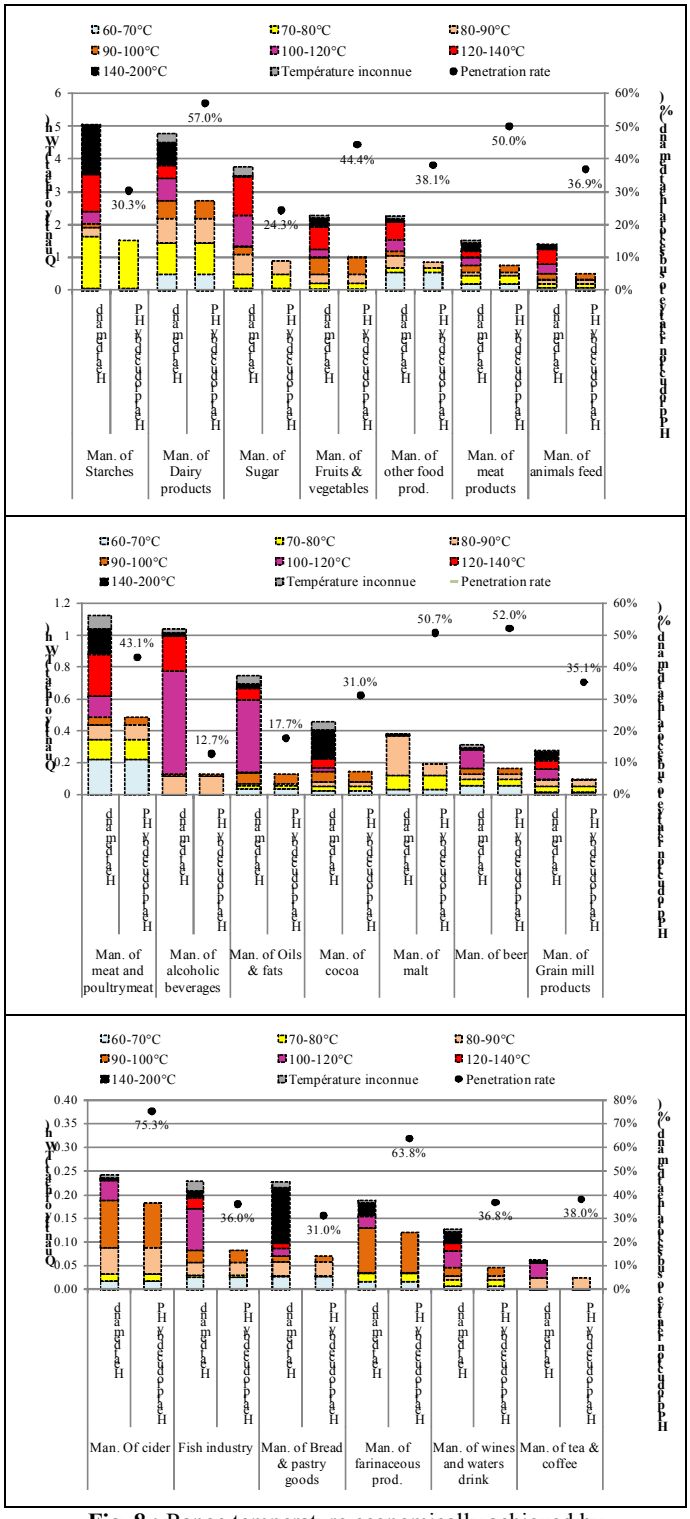



**Fig. 8 :** Range temperature economically achieved by subsectors in French F&D industry up to 2020.

We deduct that from it the level of the penetration rate of HP systems is not directly correlated to the level of the subsectoral heat demand, but rather to the most established temperature ranges, the seasonality of production and the economic environment (relation between electricity and natural gas

prices). **Furthermore, these results show that HP high temperature (beyond 100°C) is not economically feasible in F&D industry until 2020.**

- ***Evolution of CO2 emissions***

Due to the reduction in the final energy consumptions in F&D industry with introduction of HP, we also observe a decrease of $CO_2$ emissions. Fig. 9 shows the evolution of reduction emissions achieved up to 2020.

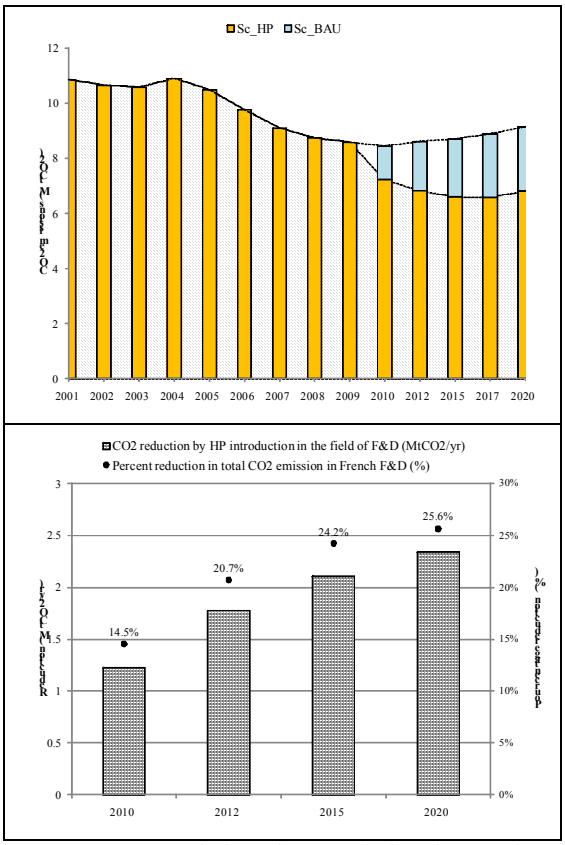



**Fig. 9 :** Impact on evolution of $CO_2$ emissions in French F&D industry.

The BAU scenario shows that CO2 emissions post(show) a decline until 2010 because of the impact of the economic recession. So, we notice a decrease of -19.5 % of emissions in 2010 compared with the level of 2005 (-5,1 % with the level of 1990). It followed by a progressive increase of emissions to reach 9,15 $MtCO_2$ due to the revival of the economic growth, that is approximately +2,8 % relative to the level of 1990). **The introduction of HP allows F&D industry to achieve a reduction of $CO_2$ emissions of around 25.6% compared with the BAU scenario in 2020, corresponding to an effort of 23.5% relative to the level of 1990**.

## 5. CONCLUSION

This implementation of heat pumps until 2020 represents around 15% of energy savings in the total final energy consumed and almost one quarter of $CO_2$ emissions avoided by F&D until 2020. A disaggregation for this energy savings and emissions avoided was done on all subsectors by temperature range and by energy end-uses, and shows a strong heterogeneity in this industry.

Furthermore, these results show that the economically feasible HP correspond to the temperature ranges below 100°C, we can aspire to reductions of $CO_2$ emissions around 35 % on the horizon 2020 compared with the level of 2005. Within the framework of the Package Energy Climate, France hopes to achieve 18.3 % of emissions reduction with sectors not covered by the EU-ETS such as those of the NEI industry, between 2005 and 2020 . This industry could contribute so effectively to the objectives fixed by action plans for the energy efficiency of the European Union thanks to the promotion of heat recovery with HP systems.

Heat pump is an excellent and very promising technology which is expected to be adopted widely in industrial sectors through further technology research and development. By applying them, it is possible to obtain value-added benefit such as reduction of $CO_2$ emissions; energy cost savings and is a pragmatic technology to fight global warming.

## 6. REFERENCES

[1] G. Seck, A methodology for modelling the Non-Energy Intensive Industry, ETSAP Workshop, Sophia-Antipolis (France), December 15-17th 2008.
http://www.iea-etsap.org/web/Sophia_Antipolis_12_2008.asp

[2] Centre d'Etudes et de Recherches Economiques sur l'Energie, A french Statistical Institue, www.ceren.fr

[3] Panorama des industries agroalimentaires, Ministère de l'Agriculture, de l'Alimentation, de la Pêche et des Affaires Rurales (MAAPAR). http://panorama-iaa.alimentation.gouv.fr

[4] U. Remme, TIMES training Workshop, Gothenburg, March 28-30th 2001.

[5] R. Loulou, U. Remme, A. Kanudia, A. Lehtila and G. Goldstein, Doumentation for TIMES model, ETSAP (www.etsap.org), April 2005.

[6] M. Gargiulo, Getting start with TIMES-VEDA, ETSAP (www.etsap.org), May 2009.

[7] M. Duminil, Théorie des machines frigorifiques: Machine à compression mécanique, Techniques de l'ingénieur B9730.

[8] J.F. Berail, E. Sapora, La pompa à chaleur industrielle à Haute temperature (PAC HT), internal report at EDF R&D.

[9] Guide Practice Guide (GPG), Heat recovery from air compressor, n°238, First published March 1998.

[10] M. Dupont, A. Sapora, The heat recovery potential in the French industry: which opportunities for heat pump systems?, ECEEE 2009 Summer study (www.eceee.org)

[11] Industrial Energy-related Technologies and Systems (IETS), Industrial excess heat recovery technologies & applications, International Energy Agency (IEA), October 25th 2010.

[12] Service de l'Observatoire et des Statistiques (SOeS), Chiffres clés du climat : France et Monde, Edition 2011. www.statistiques-developpement-durable.gouv.fr

# Innovative uses of available Science Technology Engineering and Math (STEM) infrastructures to improve motivation and retention of urban middle and high school students

Dr. Joseph Bowman, Jr.
Center for Urban Youth and Technology
University at Albany
Albany, New York, 12222, United States

and

Dr. Joseph Bowman, Jr.
ED 114B, University at Albany
1400 Washington Ave.
Albany New York 12222

## ABSTRACT

This presentation focuses around the nation's K-12 STEM initiative, concerns about our student's performance in math, science, and reading, and the "Quiet Crisis" that impacts American competitiveness and innovation. In reviewing these issues, identification of resources, planning methods, and program interventions are introduced.

A presentation of Innovative Infrastructure Use "cases studies" in higher education and state agencies are provided. Middle and high school student concerns about Motivation and Retention are discussed from their perspective. The role of the Center for Urban Youth and Technology (CUYT) has a design engine of culture-base STEM models is presented and their Institute for Nanoscale Technology and Youth (INTY) is discussed. The collaboration between CUYT and the College of Nanoscale Science and Engineering is reviewed. We make the case for collaboration, the use of existing in STEM infrastructures, and the important role the these interactions have on students motivation and retention issues.

**Key words:** Student motivation, K-12 retention, quiet crisis, collaboration, STEM, culture base, and informal education.

## 1. INTRODUCTION

President Obama's, 2010 report, "Prepare and Inspire: K-12 Science, Technology, Engineering, and Math (STEM) Education for America's Future," is a central theme of this article.

> "STEM education in K-12 relies on our ability to motivate and inspire students. This involves creating exciting opportunities for students to have individual or team-oriented experiences with the ideas, discoveries, and emerging knowledge in STEM fields. …Key avenues can provide students with experiences that allow them to explore and challenge themselves with STEM. …. Out-of-class and extended day activities that include contests, laboratory experiments, field trips, and more". (PCAST p.87.)[1]

This telling statement indicates our potential future in the field of STEM education.

However these are some sobering statistics about America's youth in the Highlights from The *Programme for International Student Assessment* (*PISA*) 2009 study on student performance in math, science, and reading.

- Average scores of 15-year-old students rank 25th out of 34 countries when compared with top students elsewhere in the world. (p. 18) [2]

- Average scores of 15-year-old students on science literacy scale, Ranked 17th out of 34 OECD countries. (P.24) [3]

- By the end of 8th grade, U.S. students are two years behind in the math being studied by peers in other countries. [4], and the Broad Foundation stated that:

- Sixty eight percent of 8th graders can't read at their grade level, and most will never catch up. [5]

There are numerous opportunities in biomedicine, nanoscience, alternative/renewable energy, e-transportation, e-literacy, e-mobile, and telecommunications, but how do we prepare our youth for this bright and exciting world of the 21st Century?

## 2. QUIET CRISIS?

The crisis stems from the gap between the nation's growing need for scientists, engineers, and other technically skilled workers, and its production of them. As the generation educated in the 1950s and 1960s prepares to retire, our colleges and universities are not graduating enough scientific and technical talent to step into research laboratories, software and other design centers, refineries, defense installations, science policy offices, manufacturing shop floors and high-tech startups. This "gap" represents a shortfall in our national scientific and technical capabilities. [6]

The notion of the "Quiet Crisis" and the impact it is having in American innovation and competitiveness must be considered and addressed. President Shirley Ann Jackson, Rensselaer Polytechnic Institute states:

> "U.S. demographics have altered, creating a "new majority," which now is comprised of women and other groups traditionally underrepresented in science and engineering. It is to this talent pool we must turn to help build the next generations of scientists and engineers, while spurring the interest of all of our young people — and herein lies our challenge. I have termed this overall situation "The Quiet Crisis" — a "perfect storm" of converging trends threatening the American innovation enterprise." [7]

The "Quiet Crisis" concerns are echoed by New York Times columnist and author of "The World is Flat", Tom Friedman and Peter Smith, Author of 2006 "The Quiet Crisis" book publication. In an interview with Susan Walsh Veronikas (Texas Tech University) and University Business, Smith States,

> *"*When you look at where we're successful, with whom we're successful and you cut that by race and income, you find that people with less money--predominantly people of color--simply are not at the table of opportunity. In numbers, they don't come anywhere close to approximating where they are in this society and where they are going in sectors of the population. We simply have to develop a national strategy for how it is that we're going to respond to those people to bring them into the mainstream of society, or there is no way we will have a middle class characterized or described by educational attainment, which is what we have agreed that we want in the next 15 to 20 years." [8]

These are a few concerns raised by "The Quiet Crisis." Included are: student scores in math and science internationally, retirement age of

the existing STEM workforce, changing demographics of our nation, college and university readiness, and inclusion of minorities, women, and adult learners into the STEM workforce.

I have identified the problem, established an audience, and provided a rationale for this project. The Center for urban Youth and Technology (CUYT) designed interventions and collaborations with existing STEM resources across New York State. We will discuss these existing infrastructures, provide case study examples of our partnerships, and present how we provided motivation and retention opportunities for urban students.

## 3. INNOVATICE INFRASTRUCTURE USES

New York Tech Valley – a region that stretches from the Canadian border near Montreal to just north of New York City – is brimming with opportunity and excitement. The area encompasses the Capital Region and parts of the Adirondacks/North Country**,** Hudson Valley and Mohawk Valley. Tech Valley offers rewarding career and business prospects, world-class educational and research facilities and a fantastic quality of life. This is the environment our youth live and are educated. It is our responsibility to expose students to the best opportunities in business, technology, education, and life-long learning.

### Higher Education - University

CUYT collaborates with other university departments (Geography and Planning, Theatre, African Studies, and Music) at the University at Albany, the College of Nanoscale Science and Engineering (CNSE), Rensselaer Polytechnic Institute, CUYT Foundation, state/city agencies, the business community, school districts, community based organizations, and cultural centers to provide program opportunities, to address these concerns.

Collaboration with atmospheric sciences and geography and urban planning department provided opportunity for students to work with the *National Oceanic and Atmospheric Administration* (NOOA) to view area weather patterns, launch weather balloons, observe weather simulations, and visit their Doppler radar facility. This experience was linked to their middle and high school science curriculum on climate, weather patterns, and the environment. In addition, students had the opportunity to do the weather forecast using green screen technology.

In Urban Planning, students surveyed their communities, identifying landmarks, parks, churches, bodegas, vacant lots and buildings, and schools. When students completed the survey of their community, they returned to the area and created a beautification plan. Lots were cleaned out, flowers were planted and vegetable gardens were seeded and watered. Using Arch View and Map Info software, the students created a community asset map. When a new school was being built in their district, the students submitted floor plans for the rooms and floor design. During their community survey, they monitored traffic flow around area schools and met with local city officials to request stop signs and lights. Finally, students documented (video and pictures) each area block and they produced a community documentary.

In the School of Education the students produced video programs, public service announcements, digital newsletters, web sites, and social network pages. Camera operation, studio production, radio production, and digital editing were all part of their on-going training activities.

The Institute of Nanoscale Technology and Youth (INTY) is a partnership between CUYT and the College of Nanoscale Science and Engineering (CNSE), (University at Albany) with a goal of engaging youth in the science, engineering, and economic impact of nanotechnology. INTY focuses on themes in nanoscience, (nanoengineering, nanobioscience, nanoeconomics, and alternative energy),

robotics, character education, radio frequency aircraft, multi media design, video production, and professional development.

INTY is an informal education (defined as: casual and continuous learning from life experiences outside organized formal or non-formal education) [9], culture-base model that provides program activities on weekends throughout the academic year and a summer residency program for middle and high school students. During the school year, INTY provides workshop sessions (Nano career, community, and high tech) at University at Albany and CNSE in Albany, New York. We visit the school districts to hold Nano Expo activities, which support teacher and parent involvement, student demonstrations and presentations of their work, and program recruitment. Student participation in the summer institute varies from one to fours weeks. The program activities are determined by availability of space, faculty scheduling, and the amount of funding. A list of program activities has been provided and radio frequency aircraft is selected for this discussion.

Students were asked to build an aircraft from a kit, cutting out the parts of the aircraft, gluing elements together (wings, rudders, ailerons, flaps, elevator, spoiler, and stabilizer), and attaching propellers, nano control devices, and the motor to prepare for flight. Students were introduced to the elements of flight (lift, drag, thrust, and propulsion) and the remote control device. Most of the day was spent on building the aircraft, testing the controls, and making sure that the aircraft were secure. The next day, depending on good weather, the students applied their skills and knowledge and flew their aircraft. We combined nano chip technology, aviation history, aeronautics, and weather to provide a STEM experience for students that have not been exposed to these fields of study.

**State Agencies**

Many of the state agencies in New York State have youth program divisions that introduce, recruit, and provide demonstrations of agencies activities and resources. The Department of Transportation (DOT) and the Department of Environmental Conservation (DEC) have been selected for this discussion.

**Department of Transportation (DOT)**

Working with the Department of Geography and Urban Planning at the University at Albany, our program was introduced to the "Bridge Building" program at DOT. DOT had a group of civil engineers, urban planners, and geologists, who supported the transportation infrastructure (bridges, highways/thruways, railway service, and construction) of New York State. The DOT program had students work in teams and we used West Point bridge building software to build bridges, and test them for road efficiency, weight, safety, structural stability, and weather resistance. The students competed against each team to determine which bridge would withstand the tests and remain standing. The engineers then met with students and created a model of the winning bridge that was created in the bridge building simulation. DOT engineers took the students to the Castleton Concrete Factory to see how bridge components were designed and created.

**Department of Environmental Conservation (DEC)**

The DEC trained our students to use Global Positioning System (GPS) scanner units, and provided GPS scanner units, so that they could visit *Five Rivers Environmental Education Center in Delmar NY to look for caches.* Called "Geocaching", this is a great adventure game for those who love both technology and the great outdoors. Students obtained coordinates of caches located in their region. The 'treasure cache' always contain a notebook or sheet on which a triumphant player can record his or her success.

DEC introduced students to alternative and renewable energy resources with demonstrations of hydro powered cars and wind turbines. Teams of students created operational wind turbines and visited the Port of Albany to see wind turbine components arrive. The students then visited a wind turbine farm in the Utica, New York area and learned about how much power the wind farm generated, the cost saving, and households/businesses covered.

We have created partnerships with the business, education, community based organizations **(**Center for Economic Growth and Workforce Investment Board), and could continue these discussions. These case examples demonstrate that these resources exist across the country and around the world, but they remain unknown unless go out and network, research the possibilities, and design STEM interventions for youth.

## 4. MOVTIATION AND RETENTION

How do we motivate and retain students with STEM to keep them in school?
Program attitudes toward students, culturally diverse content, non-threatening environments, hands-on activities, and a culture-based model (CBM) are concerns that have been integrated into this STEM initiative. We view our students as resilient.

The Programme for International Student Assessment (PISA), *Against the Odds: Disadvantaged Students Who Succeed in School* Report provides a working definition:

> "Resilient students are characterized by their positive approaches to learning or, more specifically……. increased self-confidence or interest in science. The evidence in PISA shows that positive approaches to learning tend to boost the performance of socio-economically advantaged students more than that of disadvantaged ones. Therefore, in their aim to foster positive learning approaches, policies should target disadvantaged students more than others". [10]

Attitudes toward our youth must change! It is not about the poor down trodden, hopeless, academically challenged student that someone on a white steed will ride in an save. Many of the students want us to believe in them, provide them with access to information and technology, and get out of the way so that they can shine. I cannot tell you how many times I have tripped over my feet in the way of student goals and desires. I am still learning to support, provide access, and step back. I have always created a culture-base model for all our student interventions and programs.

Dr. Patricia Young in Instructional Design Frameworks and Intercultural Models states:

> "Culture-base Model is an intercultural instructional design framework that Guides designers through the management, design, development, and assessment process, while taking into account explicit culture-based considerations". [11]

Inherent in the CUYT culture-base model are other features that impact the programs and interventions. Selection of qualified staff from diverse backgrounds and gender is a fundamental issue. Students, staff, and administration have a common background that is identifiable. Invite community, business, legislature, church, sports, science, and students' leaders as role models to meet with students. Provide cultural activities for students that celebrate their cultural heritage. Feature students completed work and accomplishments as a way to stimulate and encourage our goal of middle and high school retention. With their permission, feature all student accomplishments via various media forms. Create interventions that continue for multiple years to strengthen the relationships that have been created.

Our uses with technology provide real time experiences for students that allow them to see,

experience, and complete activities of their own design and creation. These activities and completed productions are used to stimulate deeper questions and interest about how these technologies operate and their relationships to math, science, and engineering. Creating hands-on, non-threatening learning environments for students can supplement and improve their educational achievement, advancement, and increase contributions to society.

## 5. CLOSING

Innovative uses of available infrastructures are all around us, but designers must reach out to the state, business, and federal agencies for support. Many community-based organizations are looking for partnerships with higher education institutions. Funding remains a critical issue and we must continue to advocate with our legislative leaders for support and resources.

## 6. BIBLIOGRAPHY

[1] President's Council of Advisors on Science and Technology (PCAST) (2010) **Prepare and Inspire: K-12 Science, Technology, Engineering, and Math (STEM) Education for America's Future**. PCAST is administered by the White House Office of Science and Technology Policy (OSTP). (p. 87)

[2] Fleischman, H.L., Hopstock, P.J., Pelczar, M.P., and Shelley, B.E. (2010). ***Highlights From PISA 2009: Performance of U.S. 15-Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context*** (NCES 2011-004). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. (p.18)|

[3]Fleischman, H.L., Hopstock, P.J., Pelczar, M.P., and Shelley, B.E. (2010). ***Highlights From PISA 2009: Performance of U.S. 15-Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context*** (NCES 2011-004). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. (p. 24)

[4] Schmidt, W. (2003, February 4). **"Presentation to Mathematics and Science Initiative."**

[5] Broad Foundation – **Education website**, http://www.broadeducation.org/about/crisis_stats.html

[6] Jackson, Shirley Ann. **(200 ) The Quiet Crisis: Falling Short in Producing American Scientific and Technical Talent.** Building Engineering and Science Talent (BEST), San Francisco, CA (p.1)

[7] Jackson, Shirley Ann. (2005). **The Quiet Crisis and the Future of American Competitiveness**. Presidential Symposium "The Business Case for Diversity in the Chemical Enterprise" American Chemical Society (ACS) Fall National Meeting Washington, D.C.

[8] Smith, Peter. (2006) **interview with Susan Walsh Veronikas** (Texas Tech University) and University Business, http://www.universitybusiness.com/viewarticle.aspx?articleid=478

[9] Definition provided by **Education.com** http://www.education.com/definition/informal-education/

[10] OECD (2011), ***Against the Odds: Disadvantaged Students Who Succeed in School***, OECD Publishing. *http://dx.doi.org/10.1787*.

[11] Young, P. (2009), **Instructional Design Frameworks and Intercultural Models** IGI Global Hershey, PA (p. XIV)

# Educator and Student Use and Development of Open Source Software and Resources

**Rita Hauck**
**Technology Studies Department, Fort Hays State University**
**Hays, KS 67601, USA**

**ABSTRACT**

This paper posits that educators should understand and support Open Source software and resources. The objectives of the study include 1) an introduction to the study of Open Source for educators 2) discussion of the history and definition of Open Source, 3) presentation of ways and means to promote the use and development of Open Source software and resources, 4) presentation of bona fide Open Source resources for educators and students, 5) presentation of samples of use and development of Open Source in education, 6) discussion of both pros and cons of using Open Source software and resources, and 7) a conclusion.

**Keywords**: Open Source, GNU, Linux, Redhat, social networks, critical thinking, authentic learning.

## 1. INTRODUCTION

Educators should do more to support student use and development of Open Source (OS) software and resources. Seemingly obvious reasons for the increased support for OS are promotion of creativity, authentic freedom of expression, freedom of access to educational resources, and practice in critical thinking and decision making skills. Learning to find, use, and adapt OS software can instill a confidence to learn and apply new information, and competence to build upon prior knowledge. Another advantage of OS is the opportunity to experience the benefits of collaboration with peers as well as innovators, and the gratification of applied learning. Such abilities and achievement may fly in the face of myths such as "I can't do math" roadblocks to learning. Rather, they support Physics Professor at Fort Hays State University in Kansas, Dr. Paul Adams, in his presentation about teaching science when he propelled himself on an auto mechanics "creeper" across the front of an auditorium, using a fire extinguisher as a personal rocket engine and being driven as well by his passion to help each student "become an Albert Einstein."

Educators can take responsibility and determine the computer software wave of the future rather than ride the crest of advancing, undefined, undetermined commercial products. The working definition of OS for this paper has a purposeful, functional, yet limited scope. This OS does not need a formal legal license, but rather has an informal license to define a non-fragmented future for education, a future punctuated by passion rather than greed. Open Source in this paper is transparent software code and "state of the art" educational materials that are available for others to modify and use as they choose in the learning environment.

The OS Initiative (OSI) organization founded in 1998 supports an OS Definition that has the ten rights explained in detail at opensource.com and further discussed later in this paper. The concept of OS dates back to the first sharing of recipes and other pre-personal computer "instant messaging" strategies among friends and other acquaintances in social networks.

## 2. HISTORY AND DEFINITION OF OPEN SOURCE

Rather than starting at the moment of the first such sharing, or with "broad spectrum" innovation giants like William Thomson[1] (1850s), this paper starts the history of OS at the emergence of the computer age when creative computer geniuses began sharing code, mathematical equations, ideas, and "software"– too many to mention here, except for Vannevar Bush (1930s), Claude Shannon (1937), Tim Berners-Lee (1989), Richard Stallman (1971), and Linus Benedict Torvalds (1991). The author was surprised at the unexpected depth of this topic which is much like an iceberg. The 80% that's under water is barely explored here; it is hoped that the links and references provided will lead the curious to a greater depth of understanding of OS and its implications for educators.

### History

The OS concept is as "old as the hills" but the current focus is on software and source code accessibility. In describing his work on the design of the navigation system for what he states "would now be called a cruise missile" – the Snark, from 1959 to 1961, Reed sets the stage for the history of OS, or the "Dawn of the Computer Era" as the article is called. It involves a continually growing community with passion to share and promote freedom to learn and innovate [1]. The Redhat website attributes the beginning of modern computer OS to the founding of ARPANET in 1968 [2].

The Microsoft Disk Operating System (MS-DOS) was the beginning of Microsoft, initially being an Open Source operating system. Hobbyists and computer

[1] "Steele, who worked full-time on the [Snark navigation] project, instead took his cues from the mechanical differential analyzers that had been developed in the 1930s by Vannevar Bush at MIT, based on 19th-century work by William Thomson, the first Baron Kelvin, [rather than the planned celestial navigation system]. These machines used wheels, gears, and cams to create a mechanical analogue of the equations. (A drafting compass can be thought of as a simple analog computer, programmed to solve the equation that describes a circle of a given radius.) The famous Norden bombsight used in World War II was a small, mechanical, very clever Kelvin-inspired differential analyzer" [1].

scientists in the 1970s through 1990s developed computer operating systems that had an umbrella term of disk operating system, commonly known as DOS. It's an example of how OS has led to some very important proprietary software systems. The various DOS systems were not compatible with each other but became proprietary systems such as MS-DOS, Apple DOS, Atari DOS, and Commodore DOS. Microsoft DOS basically "trumped" IBM with its DOS, although both companies both started in essentially the same place with the same basic operating system.

VisiCalc was the first spreadsheet software – author bought it along with IBM's groundbreaking 5150 PC in 1981! VisiCalc also began in 1979 as a hobbyist's "OS," as defined in this paper. Another example of OS is Open Office, a currently popular suite of applications that was developed by Sun Microsystems, apparently to check the Microsoft juggernaut, providing essentially the same functionality as Microsoft's Office suite. Particularly, in the third world, Open Office will likely be widely adopted, although its recent "donation" by Oracle (Sun Microsystems' successor) to the Apache Foundation may significantly change its future. Wikipedia is a classic OS, inviting volunteers to write, edit and augment articles.

While proprietary software and resources were evolving, so was the development of Linux-based products, and free Google products such as Google Earth, Picassa, and Google Documents. In 2001, "Microsoft's Ballmer calls Linux the biggest threat to Microsoft" [2]. In 2009, a Linux follower noted, "Microsoft is open source's biggest threat" [3]. This was the main reason that Sun Microsystems supported the development of Open Office and why Google supported Firefox as competition for Microsoft's dominant Internet browser.

As the competition among the technology giants continues, educators seek economical, reliable, and equitable software support and resources. While there now are multiple software sources, commercial companies are forming partnerships and foundations to entice schools to continue to buy their products and services. For example, the Bill & Melinda Gates Foundation has partnered with a foundation of the Pearson textbook and school technology company to develop courses rich with media that will promote the core standards adopted by 40 states for K-12 English and Math. Efrati and Gorman noted, "Since its entry into the Chinese-language world in 2000, U.S. search giant Google Inc. has struggled to balance its growth ambitions in the vast but restrictive new market while adhering to a self-held principle: 'Don't be evil'" [4]. Did Google adhere to principle by sleeping with the potential "enemy?" Will commercial school textbook media providers truly help educators and students to, as Prensky recommends, "Be able to follow one's passions as far as one's abilities allow, and as FHSU Professor Adams recommends, "become Albert Einsteins?"

The "Revolution OS" documentary video provides a robust summary of the history of GNU and the Linux kernel, and the 1980's free software movement beginning with Richard Stallman's experience at the M.I.T. Artificial Intelligence Lab in 1971. The "infancy period" of Linux was 1991 to 1993. The documentary features interviews with key OS players, including Torvalds and Stallman with commentary by Eric Raymond, author of The Cathedral and the Bazaar. Raymond notes in the documentary that he chose "Cathedral" to represent the traditional and hierarchical control versus the "Bazaar" that offers a variety of choices. Bruce Perens noted, "Netscape went open source, essentially to beat Microsoft" in 1998, naming its browser project 'Mozilla' after the original Navigator code name. The documentary shows the first office of VA Linux Systems where they moved in 1995, and explains it was the place where the term Open Source originated, to be used in lieu of the term 'free software' which might give the illusion of 'cheap' or have other misleading connotations, especially suggesting that someone should not want to use it due to inadequate marketing, development or resilience. Raymond attributes the term "Open Source" to Christine Peterson who was in the group making the decision. Apache web server, described as "the killer ap of web servers" touted to be the "must have technology" for Linux, motivating "many ISP and e-commerce companies to choose Linux over Microsoft's Windows." Redhat and Oracle also played significant roles in broadening the scope of Linux. The documentary also provides a history of the attempts and eventual success of some players to attract venture capitalist investment [5].

A Redhat history timeline from 1968 to 2009 and a Linux history with images from Business Insider both complement the documentary and this paper. Of interest from the Red Hat timeline's 2001 events is "Red Hat proposes to provide free software to every school district in the United States if Microsoft pays for the computing hardware in an alternative to Microsoft's class action lawsuit settlement plan" and from 2008, "Red Hat partners with the Georgia Institute of Technology to release worldwide study comparing and contrasting open source activity across 75 countries [2]. In 2001, "Microsoft announced its Shared Source campaign to combat the Open Source movement" [5]. On June 2, 2011, Oracle donated the Sun-developed Open Office to the Apache Foundation [6]. As of June 3, 2011, valinux.com is "Geeknet" with the "bottom line" on the web page noting, ". . . SourceForge.net, Slashdot, ThinkGeek and freshmeat are registered trademarks and Geeknet is a trademark of Geeknet, Inc."

Google's Android phone, which uses a Linux-based operating system, now dominates both the iPhone and Blackberry. HP's webOS has also done well with Linux. The Linux foundation Executive Director stated, "Apple is your worst enemy and your best friend if you're an open source guy" and, "Apple in many ways has done a lot of good things for open source and for Linux. It changed the definition of what client computing is. That has been good for Linux. Apple also has a lot of open source components within their products and tends to work very well, in some cases, with the open source community. But I'm not going to argue that they don't have a very closed system as well" [7].

In April 2011, Linux Foundation Executive Director Jim Zemlin claimed, "Two decades after Linus Torvalds developed his famous operating system kernel, the battle between Linux and Microsoft is over and Linux has won . . . They used to be our big rival, but now it's kind of like kicking a puppy. . . While Microsoft's stock has stagnated . . . open source torchbearer Red Hat has soared" [7]. This year is the 20th anniversary of Linux [8].

**Definitions**

As of the writing of this paper, the term "Open Source" has not been limited to software code. It can refer to an article or report, suggesting it is public information and not confidential, proprietary or restricted information. For example, the Open Source Center provides foreign open source intelligence at opensource.gov. This paper focuses on Open Educational Resources (OER) and Open Source (OS) and other free or "open" resources that can help educators and students, or anyone who wants to learn in today's knowledge society. Stallman stated, "think of freedom of speech, not free beer . . . freedom to change the software" [5]. For educators, it means freedom to educate, freedom to learn, innovate, and adapt.

Referred to in "Revolution OS" as the author of the Open Source definition, Perens defines OS in terms of its rights. The OS Initiative (OSI) organization founded in 1998 supports the OS Definition (OSD) that now has ten rights explained in more detail at the OSI website. The rights are 1) free distribution, 2) source code available, 3) derived works permitted, 4) integrity of The Author's Source Code, 5) No Discrimination Against Persons or Groups, 6) No Discrimination Against Fields of Endeavor, 7) Distribution of License, 8) License Must Not Be Specific to a Product, 9) License Must Not Restrict Other Software, and 10) License Must Be Technology-Neutral [5].

## 3. WAYS AND MEANS

**Stop Dabbling – Start Developing**

Our students need to be in the real world (authentic learning), to compete for and contribute to the greater good, not totally focused on market share or self-aggrandizement, but on passion for the good, to be better people and a better society – a better world – not for self-reliance or reliance on monopolies or government. Efrati and Gorman reported "Google said hundreds of Gmail users were tricked into sharing their passwords with 'bad actors' based in China [4]. One way to teach our students to pursue good or be 'good actors' is to teach them to work with others who are trustworthy to improve resources and provide equitable access.

Prensky notes two large issues for educators: "1) What our students need to learn for the future is, to an enormous degree, different than what we are teaching now, and 2) Future education is a worldwide issue, not just a state or national one." He presents five essential 21st century metaskills with the goal: "To Be Able to Follow One's Passions as Far as One's Abilities Allow." The five skills individuals must master are 1) Figuring out the right thing to do, 2) Getting it done, 3) Doing it with others, 4) Doing it creatively, and 5) Constantly doing it better. For a list of sub-skills under each, refer to pages 186-187 of Teaching Digital Natives [9]. All of these aspirations are completely consistent with educator support of OS.

Finding the right commercial product or OS resource is like shopping at a local supermarket – either grab the first thing that satisfies the desire, or look at the unit prices. Educators must learn more to be a better and more informed consumers. According to Harvey, "The educational community has discovered open source tools in a big way. Analysts predict that schools will spend up to $489.9 million on support and services for open source software by 2012, and that only includes charges related to operating systems and learning management systems. Teachers, professors, and home schoolers are using open source applications as part of their educational curriculum" [10].

There seems to be a tendency for educators and entrepreneurs to want to be the first to inform others about "snake oil" that really works, something that will transform education as we know it, a tendency to jump on the band-wagon of Web 2.0, Open Source, and "You-name-it!" Contrary to popular opinion, the term "Web 2.0" was first coined by Darcy DiNucci in her three-page article titled "Fragmented Future" in which she described the future of the Web as "the ether through which interactivity happens" [11]. Tim O'Reilly used the term to market conferences using the term Web 2.0. Like the term Web 2.0, the term Open Source has become a buzz word in education communities without much reflection or understanding about the history behind the terminology and concepts. Next, we'll be hearing more about the "cloud computing era," and the "cloud-age" referenced by Levy. In his book, In the Plex, Levy describes a "googler" and the Google mission: "We want to make you happy; if we can do it for free, so much the better. You just start editing in the cloud – no files, files are so 1990" [12].

Wenk notes that the OAIster database provides a central source for OER repositories. Wenk recommends that repositories provide metadata in formats compatible with most search engines indexes: Unicode for text, TIFF (Tagged Image File Format) for images, MP3 (MPEG-1, Layer 3), and MPEG-4 for video. Wenk describes Portable Document Format (PDF) for documents with text, graphics, audio, and video, but notes that "PDF is not a suitable format for reusable learning objects." Wenk notes that it is easy to adapt HTML pages and suggests a free Web authoring tool such as KompoZer [13].

## 4. OPEN EDUCATION RESOURCES

Upon questioning a colleague about where to begin research for this paper, he exclaimed, "It would amount to academic malpractice not to use Google Search and Google Scholar!" WorldCat is also good but the researcher may have to wait ten days to two weeks to

receive the item found. A graduate student in one of my past Multimedia Applications courses engaged our class in a tour of China and its culture using Google Earth. It was "almost like being there." Levy describes that Google Earth, formerly "Keynote" cost $1,000 for an annual subscription. Google purchased it and made it available for free [12].

A search at Multimedia Educational Resource for Learning and Online Teaching (MERLOT) at merlot.org unearthed a tutorial "XHTML, HTML 5, and CSS," that this author used successfully in teaching two semesters of a graduate course, "Hypermedia/Hypertext Applications," for FHSU's Masters in Instructional Technology program. The "Welcome to Merlot" piece advocates "Putting Educational Innovations Into Practice – Find peer reviewed online teaching and learning materials. Share advice and expertise about education with expert colleagues. Be recognized for your contributions to quality education." The OS contributions of Dr. Kevin S. Floyd to the MERLOT community are gratefully recognized [14]. It is truly good news that there are too many OER resources such as olcos.org and nasa.org to mention in this paper.

"Free" is seldom really free, especially with online tracking software watching our every Internet move. It isn't free to even visit some websites, as the visitor will likely end up getting targeted email advertisement messages from them. There is a huge difference in the two phrases, "free for all" and "free-for-all" The first seems self explanatory as a website may be free for all to access. A glance at the deluge of incoming email reveals the latter as described by Merriam-Webster as "a competition, dispute, or fight open to all comers and usually with no rules : brawl; also : a chaotic situation resembling a *free-for-all* . . ." [15]. To avoid the latter in finding the right OS resources, educators need the collaboration with others of similar interests, passions, and values.

Wenk notes, "The OER movement aims to break down barriers – such as teaching material locked up behind passwords within proprietary systems or filed in personal drawers – and enable sharing content freely" [13]. A visit to oercommons.org reveals their license: "Except where otherwise noted, content created on this site is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License."

## 5. SAMPLES OF OPEN SOURCE IN EDUCATION

In their article, "Dilemmas in a General Theory of Planning," Rittel and Weber substantiate that social endeavors are most often "wicked problems" [16]. Prensky has a chapter in his book that describes 132 teaching tools from "3D (Three-Dimensional) Printers" to "YouTube" with how they can be used by students and teachers, including the types of skills students may learn from each. He reminds teachers that it is not possible to know the ins-and-outs of all 132 listed tools and the many more that are not listed, or those that are emerging. He describes methods and strategies of students and teachers partnering in learning [9].

Many educators are using *mixed IT*, a term used by Microsoft's Jean Paoli in an article titled "We love open source" [17]. Kougianos, Mohanty, and Patra describe how they use OS and free software in their Digital VLSI courses in Electrical and Computer Engineering [18]. VLSI, the acronym for "very-large-scale integration," involves the process of combining thousands of transistors into integrated circuits [19]. For more information, refer to their article, "Computer Engineering Through Open-Source/Free Tools" [18]. As mentioned, the Digital VLSI courses use LDAP. In addition, they use XCircuit programmable schematic entry tool, SPICE standard analog simulator. GTKWave graphical waveform viewer. They use Magic, and graal, although Toped and LayoutEditor are also recommended. They note "the Alliance VLSI system is part of many Linux distributions and comes in precompiled binary packages, as well as complete source code" [18].

Wenk describes a successful OER search for a tutorial about the OS course management system called Moodle [13]. Drupal is another widely used content management system with samples at the Drupal.com website. The Free Software Foundation lists available engineering education software that includes AeroCalc, GDC, Icarus Verilog, IntelHEX, NumPy, and OpenVulture.

For an undergraduate "Technology in Society" course, we keep track of current news articles at Evernote. Students and I access news in our face-to-face class and "in the cloud" [20]. I also use Dropbox.com for sharing files in the cloud with students, colleagues, families, and friends. Websites such as Evernote and Dropbox can provide a sort of "information OS" – like recipe sharing, a way to experience the cloud for personal and professional purposes – and stay "ahead of the curve."

It is smart to use google docs to learn about spreadsheets. There is not a steep learning curve to transition from Google's OS cloud application to the other proprietary spreadsheet applications like MS Excel. If the OS application continues to be a robust resource, there is really no reason to "step up" to a proprietary product. Encourage students to develop more OS applications and enhancements to current applications. Especially, encourage computer science classes to contribute to developing OS software. It's a win-win "open-open" accomplishment for the greater good of education.
I teach XHMTL (because it's free!) and it engages the brain. I also let my students veer off in other directions half-way through the course. By that time, they are better able to choose another product since it seems there are more every day.

## 6. PROS AND CONS

Problems happen whether code is OS or commercial: "The recent data breach involving Sony Corp.'s online videogame services has knocked more than 6% off the Japanese electronics maker's shares, with some analysts

estimating the incident could cost the company over a billion dollars as it takes steps to soothe and protect customers . . . Sony acknowledged personal information for 77 million customers of the network had been stolen. . . . the company said hackers may have obtained personal data for customers from another online service, bringing the potential number of compromised accounts to more than 100 million" [21]. Apple is currently in a "whack-the-mole" contest with the first credible large-scale malware attack on its "invulnerable" "OS X" OS [22].

**Pros**

Open Source could be an incubator and catalyst for improvements and advancements in software by users, and enrich the student learning experience. Innovation, teamwork, and collaboration are some of the hallmarks of OS. Linux Desktop noted the vulnerability of Microsoft and any other commercial company when they each determine to remain monopolies and to ". . . kill its competitor instead of innovating . . ." [3] Google has to have its own "skunk works" because the company is so "mega open source," in order not to have its engineers bogged down with the mainstream OS technologies that Google distributes. The momentum of being a profit-driven large enterprise can inhibit creativity, hopefully the reason that Open Office has been released by Oracle to the hacker community.

San Francisco State University, working jointly with University of Applied Sciences, Fulda, Germany, and Florida Atlantic University is developing an OS Web-based electronic Teamwork Assessment Tool, *e-TAT. "*Upon successful testing e-TAT will be available as Open Source tool" to provide Teamwork and "Soft Skills" assessment in Software Engineering Education [23]. The worldwide collaborative community is certainly an advantage of OS. In Sri Lanka, OS online survey software has been translated into the local Sinhala language with efforts to develop it in the Tamil language [24]. Open Source frequently allows for a broader application of creativity in meeting local needs.

Kougianos *et al.* note the limitations of academic versions of software as being too restrictive. Open Source allows users to do more and even allows students and instructors to use their products in a creative and entrepreneurial spirit following the GNU General Public License (GPL). They prefer OS to free software because of the viability and availability of the source code in OS. They use the Lightweight Directory Access Protocol (LDAP) based on its simplicity and availability. They note some advantages of the OS Alliance VLSI system that comes with CMOS libraries and other resources and "has been fieldproven with industrial designs," is "well-established" and "allows for full customization." They conclude that their "overall objective is to demonstrate that high quality undergraduate education in VLSI is possible with open source or free tools and should be encouraged as part of the ethics training that is expected of modern engineering curricula" [18].

**Cons**

Sometimes developing or using OS may not be practical. It may also end up like a free-for-all at times. In reference to the Snark project, Reed commented that he "nonetheless realized that an enormous amount of practical work had to be done for anyone to actually use it" [1].
Some people get emotionally involved with their work. The passion of anger and hatred against Microsoft or OS makes some people or entities seem villainous. There have been accusations or assumptions that the OS and free software movements have been anti-capitalism and pro-communism, with Microsoft CEO Balmer calling OS a cancer and intellectual property destroyer [2]. Microsoft has a web site dedicated to FAQ about their stand on OS. Search, for some "PR" entertainment.

Trustability is a concern. The Software Composition Group at the University of Bern, Switzerland, is developing a "trustability" metric based on a collaborative filtering system of users and their votes relative to past contributions of developers to "trustable" projects. Gysin and Kuhn note that code search engines such as Google Code Search, Krugle, and Koders do not provide sufficient trustability information about projects returned by a search. Gysin and Kuhn used Ohloh (ohloh.net) to collect metadata for development of their trustability metric prototype JBender [25].

In their conclusion about progress on the WattDepot project, Brewer and Johnson note, ". . . we want to explore the issues associated with privacy of energy data. . . We recognize that this privacy model is quite limited, and in fact all our applications to date have used public sources, allowing their data to be freely accessed. We want to investigate more sophisticated and useful privacy models in future" [26].

Philips quotes Anderson from Free: The Future of a Radical Price: "Free is not enough . . . It has to be matched with paid. . . . Today's Web entrepreneurs have to invent not just products that people love, but also those that they will pay for. Free may be the best price, but it can't be the only one" [27].

## 7. CONCLUSION

As mentioned previously in this paper, I was surprised at the unexpected depth of this topic which I now understand to be much like an iceberg. While I caution my own students not to use too many quotations, as I continued to find timely material and learn as the paper was under development, it seemed appropriate to include such material for interested readers.

The advantages of using OS seem to far outweigh the disadvantages. Some of the advantages include increased and worldwide collaboration, learning with application and purpose, enriched learning experiences, creativity, affordability, simplicity, full customization opportunities, ethical purpose, and accommodation for local needs. The disadvantages seem to fall equally on both OS and

commercial products: Sometimes pursuing a project may seem impractical. Often negative emotions or "winner take all" attitudes and confrontations can obstruct the vision and mission. Lack of ethical behavior, trustability and privacy issues face both disciplines. The egos, missions, and passions of key innovators, educators, foundations, and companies will have an enormous impact on the future of education and ultimately, on our society.

We must keep the freedom to choose between rigidly commercial software products, however masked or denominated, and true OS applications. If everyone has to own or use the same product, creativity will be stifled. The first WordPerfect was written by faculty and students at Brigham Young University to help orchestrate the marching band performances at BYU. To bring up the Wikipedia definition or information about WordPerfect, it took a Google search of "wordperfect wikipedia" as a google of "wordperfect" brought up Corel and the interest or hope of still selling the product (four or five hits before Wikipedia). Products will have loyalists or people who don't want to try something new or different, or use OS.
There is much money to be made from education and there are so many hooks designed with a purpose of keeping students as customers forever. Computer hardware and software are completely entangled in the education process and an integral part of the education system. Open source and OER alternatives to commercial products can be a viable way to make necessary products available for faculty and students, to have the best and widest variety, and be a bazaar of educational materials and resources available to all students. Reed noted that "the Greek word for helmsman is *kybernetes*, the source of the word cybernetics" [1]. Open Source developers, and OS users as well as commercial enterprise leaders are like helmsmen at the brink of innovation.

## 8. REFERENCES


[1] Reed, I. "The Dawn of the Computer Age." ***Engineering & Science,*** No. I, p. 7-12, Article condensed from chapter three of his autobiography, Alaska to Algorithms. Xlibris corporation, 2005. http://eands.caltech.edu/articles/LXIX1/reed.pdf

[2] http://www.redhat.com/about/companyprofile/history/

[3] CoolBreezeOne**.** Microsoft's Cold War Against Linux. *Linux Desktop News.* March 1, 2009. Retrieved June 4, 2011 http://linuxdesktopnews.blogspot.com/2009/03/microsoft-sues-linux-vendor.html.

[4] Efrati A and Gorman S. "Google Mail Hack Blamed on China" **Wall Street Journal**. Retrieved June 1, 2011 http://online.wsj.com/article/SB10001424052702303657404576359770243517568.html

[5] A J.T.S. Moore FILM. "Revolution OS (History of Linux/Open Source documentary, 86m)" Wonderview Productions, 2001. Viewed May 31, 2011 at Sift Partners Inc. http://videosift.com/video/Revolution-OS-history-of-LinuxOpen-Source-documentary86m

[6] Scott J. "The software giant 'donates' the open source office suite to the Apache Foundation." Dennis Publishing Limited licensed by Felden. http://www.itpro.co.uk/633907/oracle-ditches-open-office

[7] Brodkin J. "Bashing Microsoft 'like kicking a puppy,' says Linux Foundation chief" **Network World, Inc.**, April 05, 2011. www.networkworld.com/news/2011/040511-linux-vs-microsoft.html

[8] Love, D. "From The Dorm Room To Your Smartphone: The Incredible 20-Year Story of Linux," **Business Insider** 2011. http://www.businessinsider.com/history-of-linux-2011-4#2011-continued-success-11#ixzz1OLCPgrLa

[9] Prensky, M., **Teaching Digital Natives, Partnering for Real Learning**, Corwin Press, 2010.

[10] Harvey C. "50 Open Source Tools That Replace Popular Education Apps," Antioch University, Santa Barbara, 2010. http://antiochsb.libguides.com/content.php?pid=15351&sid=1073099

[11] DiNucci, D., "*Fragmented Future,*" Print 53 (4): 32. Carla DiNucci, 1999. http://www.cdinucci.com

[12] Levy, S. **In the Plex – How google Thinks, Works, and Shapes Our Lives**, New York, Simon & Schuster, 2011.

[13] Wenk B., "Open educational resources (OER) inspire teaching and learning," *IEEE EDUCON Education Engineering 2010 – The Future of Global Learning Engineering Education, April 14-16, 2010, Madrid, Spain,* p. 435-441.

[14] Floyd K. "XHTML, HTML 5, and CSS Tutorial," 5th edition, Macon State College, 2010. http://www.maconstateit.net/tutorials/XHTML/default.php

[15] www.merriam-webster.com/dictionary/free-for-all

[16] Rittel H., Webber, M. "Dilemmas in a General Theory of Planning," ***Policy Sciences*** 4 (1973), 155-169. © Elsevier Scientific Publishing Company. http://www.uctc.net/mwebber/Rittel+Webber+Dilemmas+General_Theory_of_Planning.pdf

[17] Brodkin, J. "Microsoft: 'We love open source' No longer calling Linux a 'cancer,' Microsoft open source strategy continues to shift" **Network World**, August 23, 2010. http://www.networkworld.com/news/2010/082310-microsoft-open-source.html

[18] Kougianos, E. Mohanty, S., Patra, P., "Computer Engineering Through Open-Source/Free Tools" NanoSystem Design Lab. (NSDL), Univ. of North Texas, Denton, TX, USA.

[19] en.wikipedia.org.very-large-scale

[20] http://www.evernote.com/pub/hauck/technologyinsociety

[21] Osawa J. "As Sony Counts Hacking Costs, Analysts See Billion-Dollar Repair Bill" **The Wall Street Journal**. Asia Technology, May 8, 2011.

[22] O'Grady J. "Apple plays Whack-A-Mole with malware authors," ZDNet, **CBS Interactive**, 2011. http://www.zdnet.com/blog/apple/apple-plays-whack-a-mole-with-malware-authors-updated/10239?tag=nl.e540

[23] Petkovic, D., Thompson, G., Todtenhoefer, R., Huang, S., Levine, B., Parab, S., Singh, G., Soni, R., Shrestha, S., "Work in Progress – e-TAT: Online Tool for Teamwork and 'Soft Skills' Assessment in Software Engineering Education. Frontiers in Education Conference (FIE), 2010 IEEE, 27-30 Oct. 2010, S1G.

[24] Jayasundara, B., Wickramasuriya K, Shakila L., "Localisation of the LimeSurvey Software," Proceedings of the Conference on Localised Systems and Applications 2010.

[25] Gysin F., Kuhn A. "A Trustability Metric for Code Search based on Developer Karma," Software Composition Group, Feb. 25, 2010.

[26] Brewer R. and Johnson, P. "WattDepot: An open source software ecosystem for enterprise-scale energy data collection, storage, analysis, and visualization," First IEEE International Conference, November 4, 2010.

[27] Philips J." To Rake It In, Give It Away: Can a business thrive that doesn't charge for most of its goods and services? **The Wall Street Journal**. Retrieved June 4, 2011 from http://online.wsj.com/article/SB124701229573408977.html#

# URM GeoPortal4Everybody: the modern platform for vocational education

**Karel JANECKA**
**Faculty of Applied Sciences, University of West Bohemia**
**Pilsen, Czech Republic**

**Karel CHARVAT**
**Help Service Remote Sensing**
**Benesov, Czech Republic**

**and**

**Raitis BERZINS**
**Institute of Mathematics and Computer Science, University of Latvia**
**Riga, Latvia**

## ABSTRACT

This paper presents a new concept for a modern educational platform. The proposed platform for vocational e-learning is designed and implemented as a virtual database based on social networks principles. It integrates different technologies like geographic information system, e-learning, multimedia and virtual reality. In particular the solution is built upon the GeoPortal4Everybody principles. The GeoPortal4Everybody is a set of modules and services which are able to communicate through interoperable services. The solution is modular and could be easy modified for different purposes. The key part of the proposed platform is the BizBiz tool. It is a web browser based e-conference collaboration and learning tool which allows conference members to watch live video enabled presentations remotely from their computers. For content management the SimpleCMS technology is used. Main advantage in comparison with other content management systems is simple approach for solving complex tasks where on the code side the clarity and security of the implementation is the main target. There already exist several implementations of our new platform around the Europe mostly in the field of spatial planning.

**Keywords:** Educational Platform, Video Lectures, Uniform Resource Management, Content Management System, Metadata

## 1. INTRODUCTION

An educational process must reflect a rapid development in the area of information and communication technologies. The using of these newest technologies offers a modern and effective education. We propose a new solution for vocational education based on Uniform Resource Management (URM) principle. Our solution offers a classical teaching (classes) and distance education as well. The emphasis is put on content management. This content in practice consists of documents of various types. The solution we propose in this article also offers a possibility to effective search through metadata records. On principle the modern platform for vocational education must be composed from the tool for education containing functionality for video lectures, drawing tools, whiteboard, management of permissions etc. Furthermore, the platform should enable searching through metadata and also enable web 2.0 functionality.

The rest of the article is structured as follows. Section 2 introduces the basic principles and components of our new technological platform for vocational education. The BizBiz tool as a learning video tool and its main functionality is described in section 3. The description of the next component of educational platform, GeoPortal4everybody and principles of Uniform Resource Management are stated in section 4. The content management and searching in related metadata are described in sections 5 and 6. Conclusions are contained in the last section 7.

## 2. PLATFORM FOR E-LEARNING

### Web 2.0

The concept of "Web 2.0" began with a conference brainstorming session between O'Reilly and MediaLive International [1]. There are many definitions of what Web 2.0 is. For the purpose of this report we have selected the

following one: “Web 2.0 is the network as platform, spanning all connected devices; Web 2.0 applications are those that make the most of the intrinsic advantages of that platform: delivering software as a continually-updated service that gets better the more people use it, consuming and remixing data from multiple sources, including individual users, while providing their own data and services in a form that allows remixing by others, creating network effects through an "architecture of participation," and going beyond the page metaphor of Web 1.0 to deliver rich user experiences.” [1]

Web 2.0 can be described according to the report by [2] and research of O’Reilly and Forrester. Web 2.0 is composed of a set of:

- **Technologies** - building blocks of web 2.0, e.g. Ajax, XML, Open API, Microformats, Flash/Flex and other techniques for creation of web applications.
- **Applications** - allow easy publishing, information sharing and collaboration. They include blogs, wikis, podcasts, RSS feeds, tagging, social network sites (e.g. Facebook, MySpace), search engines, Massive Multiplayer Online and others.
- **Values -** they build on the knowledge and skills of the user:
    - user as a content producer/provider,
    - user providing feedback, comments, reviews of the existing content,
    - users who access, read and watch the existing content,
    - user who does not benefit from web 2.0 applications and values.

**The new concept for educational platform**

The proposed platform for vocational e-learning is designed and implemented as a virtual database based on social networks principles of the GeoPortal4everybody (next also as a “geoportal”) and principle of web services using URM [3]. It integrates different technologies like GIS, e-learning, multimedia, and virtual reality. Important part is integration of social networking tools. These services are not implemented on geoportal directly but are implemented as virtual services on different places in the world. The access is guaranteed through a single access point. Single access point is based on GeoPortal4everybody bus component and SimpleCMS Social Network Content Manager. Bus includes single authorization and interconnection of components as illustrated on figure 1.

The tools implemented in the training platform for education are as follows:

- BizBiz,
- GeoPortal4Everybody based on the Uniform Resource Management principles,
- SimpleCMS.

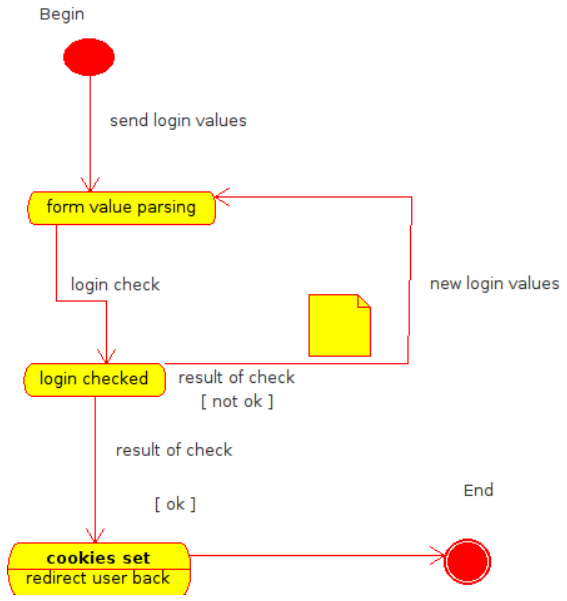



**Fig. 1.** Authorization module supports authorization into all components of the educational platform except BizBiz.

## 3. BIZBIZ TOOL

The key part of the educational platform is the BizBiz tool [3]. BizBiz is our own learning video tool for training platform. It is integrated into GeoPortal4Everybody solution. BizBiz is a web browser based e-conference collaboration and learning tool which allows conference members to watch live video enabled presentations remotely from their computers. It does not require users to install special software apart from modern web browser and java virtual environment and most of the configuration connected with webcam and screencast is done automatically. Additional equipment is headphones, optionally webcam. BizBiz focuses on providing a learning platform for spatial planning and GIS but can be used also in other applications. BizBiz is free of charge and published under open source (Affero GPL) licence in source forge.

In its foundation BizBiz is an open system, which focuses on the ease use and low costs for the users as well as system maintainers. It uses popular standards and formats such as HTML5 and Flash where unavoidable. The downside of this system is reduced quality of video and audio when compared to other telepresence and embedded systems, since the Real-Time Messaging Protocol (RTMP) protocol uses unicast addressing methodologies where every video stream is transmitted separately thus having an increased network bandwidth usage. Adobe Flash video codec support for live video broadcasting is also limited, but it is sufficient for most usage scenarios, which previous usage feedback shows. The system's network traffic has been modelled using discrete event simulation as well as statistical and

analytical approaches and total bandwidth predictions for different usage scenarios have been made.

The main features of the BizBiz tool are as follows:

- Multilingual – BizBiz currently supports English, Czech, Bulgarian, Italian, Greek, Latvian, Lithuanian language, but more languages could be added if necessary,
- lectures creating and editing – BizBiz supports tools for preparing and editing lectures by lecturers,
- lecturers live narration using web camera provided video and audio,
- Whiteboard (WB) - WB allows to put on it following elements:
    - images (e.g., slideshows) presentation,
    - simple screencast using an automatically installed Java application,
    - webpage demonstration through screen casting,
    - online maps,
    - synchronous pre-recorded video display,
- drawing tools for drawing on WB elements,
- keyword list and glossary of lecture,
- uploading files necessary for lecture,
- metadata description of lecture,
- mechanism of permissions – the lecturer is able to grant following permissions to individual student:
    - broadcast video,
    - broadcast audio,
    - upload files,
    - slide transition,
    - draw on slides,
    - edit lecture data,
    - edit slides,
    - edit glossary,
    - switch groups,
- text chat,
- polls and surveys – it allows the lecturer to conduct questions with multiple choice answers directed to the audience,
- student test mode – each student in separate sandbox. Lecturer can join each one and monitor their work lecture recording,
- lecture playback.

The training process can be realized as:

- online video lecture,
- recorded lecture,
- publicly available resources.

Online video lecture is very effective method. But there is one shortcoming – there is fixed date and time when online video lecturing is being conducted. Therefore there may be situations when many potential audience members cannot participate. Recorded lectures enable watching the lecture in acceptable time for every interested person. Therefore recorded lectures also play important role in training.

Every lecture's life cycle consists of several stages:

- upcoming lecture – when a lecture has been created by lecturer and published in BizBiz lecture list,
- live lecture – when a lecture is online,
- archived lecture – after end of online lecture.

**Upcoming Lecture**

This is the stage where a lecturer creates new lecture. After this creation the lecture data could be modified till next stage. Lecturer submits the following data for every new lecture:

- name, surname,
- short description of lecture,
- lecture access (public, with invitation, with password),
- planned start time,
- keywords,
- glossary,
- uploads necessary files,
- submits data for WB: uploads images (presentation slides in ppt, odp, jpg, png, gif formats), web links, maps and uploads pre-recorded video files or inserts embeddable videos from popular video sites like YouTube,
- metadata about lecture in Dublin Core metadata standard.

The lecture is published for audience who use BizBiz. After publishing students are able to look all data about upcoming lecture and prepare to it. They can send their suggestions and notes by text chat to the lecturer and the lecturer can make modifications.

**Live Lecture**

The lecture is online. There are several ways how the lecturer can manage the lecture.

***Simple lecture without feedback***. In this case the lecturer takes full part in this event – he/she uses webcam, shows slides on WB, draw on slides, shows maps and pre-recorded video files. Audience mainly have means to watch and hear the lecturer on their personal computers and contact with the lecturer and others by chatting.

***Lecture with feedback***. There are many possibilities:

- The lecturer can allow video browsing to some students. It can be done by mechanism of permissions. In this way the lecturer has feedback with those students whom he allowed video browsing. Then after little dialog between the lecturer and those students, the lecturer can switch off these permissions and switch on for other

student/students. Thus the lecturer can organise feedback with audience.

- The lecturer can achieve feedback with audience by switching on permission draw on slides to one or more student. It can be used when the lecturer shows maps on WB.
- The lecturer can use BizBiz feature polls and surveys. It allows the lecturer to make polling about interesting questions.
- Using BizBiz feature student test mode the lecturer can separate students in separate sandbox. The lecturer can join each one and monitor their work lecture recording. For example each student can draw the solution of some exercise on a blank slide and only he and the lecturer can see it.

**Archived Lecture**

This is a stage when the lecture is ended by the lecturer. After that, no user can join the lecture, until it is started again. Each lecture after ending can be prepared by lecturer for playback. Audience can playback them individually. What about lectures which didn't start. There are several possibilities:

- The lecturer can edit lecture's data. If planned time is modified then the life cycle of lecture changes and it is upcoming lecture stage.
- Audience is able to look all data about this lecture.
- The lecturer can delete this lecture.

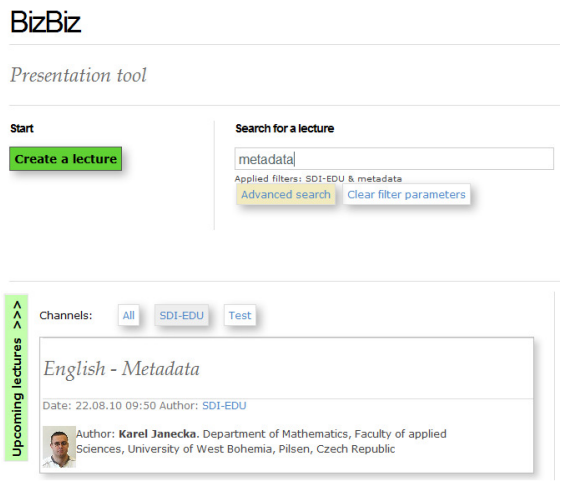



**Fig. 2.** An example of finding of a relevant lecture using metadata.

## 4. GEOPORTAL4EVERYBODY & UNIFORM RESOURCE MANAGEMENT

Uniform Resource Management provides a framework in which communities can share information and knowledge through their description, which is easy understandable inside of the community [3]. In order to effectively share information and knowledge, there is a standardized schema, which supports uniform description of information and knowledge including common vocabularies. A schema defines the meaning, characteristics, and relationships of a set of properties, and this may include constraints on potential values and the inheritance of properties from other schemas. The schema specification language is a declarative representation language influenced by ideas from knowledge representation (e.g. semantic nets, frames, predicate logic) as well as database schema specification languages and graphic data models. The context characterizes any information, knowledge and observation. Context strongly influences the way how the information will be used. There exist different definitions of context.

The important issues for the context are:

- an identity of an entity,
- a profile of an entity,
- a spatial information,
- a temporal information,
- an environmental information,
- a social relation,
- resources that are nearby,
- an availability of resources.

The GeoPortal4Everybody is a new, integrated solution being designed as combination of previous technologies - Uniform Resource Management, Geohosting [4] and new technological development of a visualization client based on HSLayers [4].

GeoPortal4Everybody is a place which allows users to search, view, examine and share spatial and non-spatial data. GeoPortal4Everybody is based on interoperability standards (OGC, W3C, OASIS, ISO) which are connected to other resources on web and help to create distributed structure of information and knowledge based on spatial localisation. Geoportal should not be closed central storage of spatial data without possibility of redistribution of this data. Geoportal should not be a solution that doesn't support searching of data and information and their viewing and using by external sources.

GeoPortal4Everybody is a concept, which is based on the next principles [5]:

- independent components,
- composition according to user requirements,
- based on Service Oriented Architecture (SOA),
- possibility to integrate with other resources,
- maximum openness,
- open source,
- open standards,
- extension to non-GIS community,
- open search,

- administration of other (non-spatial) data sources.

The GeoPortal4Everybody is not one integrated solution but a set of modules and services which are able to communicate through interoperable services (OGC, W3C). The solution is modular and could be easy modified for different purposes. The GeoPortal4Everybody is based on Open Source technologies, but it could be integrated with different technologies like MS SQL or ArcSDE. URM supports validation, discovery and access to heterogeneous information and knowledge. It is based on utilization of metadata schemas. The URM models currently also integrate different tools, which support sharing of knowledge. The geoportal contains common visualization, data sharing, metadata and catalogue functionalities. Additional parts of solution could be also tools for management sensor observation and spatial data transformation and processing.

The GeoPortal4Everybody contains these parts:

- authorization,
- Simple Content Management System – Simple CMS,
- catalogue client,
- visualization client,
- metadata editor,
- Geohosting
    - MapMan
    - DataMan
- Metadata Extractor tool.

The applications are working relatively independently; main communication is done through metadata system. The detail description of all above mention components can be found in [4].

## 5. SIMPLECMS FOR CONTENT MANAGEMENT

In our educational concept we prefer system of distributed data sources where every provider could decide about accessibility of his data against concept of cloud computing, where one organisation is managing all information. The idea is not against outsourcing of data or services or using of external services but to have control about data from the side of data holders. For implementation of this concept we suggest ideas of GeoPortal4Everybody.

System SimpleCMS - Content Management System (SCMS) is focused on usability and simplicity for end users on their mind [3]. Main advantage in comparison with other CMS systems is simple approach for solving complex tasks, where on the code side the clarity and security of the implementation is the main target. The using of SimpleCMS Content Management allows exploiting the potential coming from social networks.

Simple CMS allows editing of home page of geoportal which is built upon the GeoPortal4Everybody principles. This function is dedicated to administrator. SCMS allows the following:

- define content and system of menu for home page,
- publish articles on home page,
- publish external links in menu on home page,
- publish predefined map composition from MapMan on home page,
- order information on home page,
- remove information from home page,
- publish RSS channels on home page.

### Menu

User can define any menus and submenus. Any menu or submenu can be external link (link pointing to any place on WWW and not into the CMS itself), where the redirected functionality is implemented and users can return back to the CMS using visible controls. It supports inserting different web applications into web pages for training purposes. Any menu can be set as Homepage, where of course only one Homepage per SimpleCMS instance is possible. Menu ranking can be reordered in any way to best fit any updates that might be required during usage.

### Article

Content holders are holding the data we want to publish. As regular articles they are composed from the “perex” and the content itself. Using nice WYSIWYG editor provides nice user experience to beginners. Support for full inline html can please any person wanting to do more fine grained look of desired article. Editor allows inserting of multimedia content like videos, photos, etc. There is a special support for inserting of dynamic maps, presentation from SlideShare, YouTube and other social networks for sharing content. Each article can be enhanced by adding various file attachments, which will be described later.

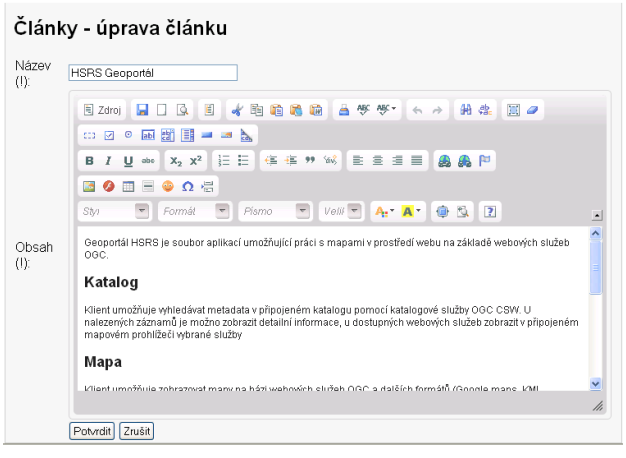



**Fig. 3.** Editing of article using SimpleCMS

**Message**
Each menu can be accompanied by the message item, which in reality is simplified article that contains just one view on short text. That does not implement detailed view so it could be described as "perex" only. Since the message(s) is always on top over the articles they can be best used as the menu description. Also here is used the some WISIWIG editor, with full functionality like in articles.

**RSS**
The CMS supports including any RSS feeds from remote sites. This allows nice and handy way of promotion for our/friendly services that we want our consumer to know about.

**Translations**
The controls of the CMS can be translated into various languages where the gettext localisation system is used, which provides easy translations to most known languages and possibility to use already created vocabularies from any other open-source project, which makes translating work really simple.

**Remote articles promotion**
Each menu in the CMS setup has automatically generated RSS feed for grabs to others. Also optional connectivity to Posterous is possible, so your web presentation can be presented on all various social networking sites (for example Facebook).

## 6. SEARCHING IN CONTENT USING METADATA

The possibility to search a valid content via metadata is a crucial part of a modern educational platform. We proposed our own tool – Metadata Extractor which allows to get metadata and process them in effective way.

**Metadata Extractor**
Metadata extractor is a tool to extract available metadata directly from different files (documents, presentation, etc.), edit this metadata and publish metadata and files on geoportal based on GeoPortal4Everybody principles. Other possibility is to extract metadata (and then edit) directly from existing URL addresses and store metadata on URM portal. Access to information is then trough direct URL addresses. Currently metadata extractor supports:

- publishing documents on the geoportal (based on GeoPortal4Everybody) – you can select any type of file in your computer, extract and edit metadata and publish this file on geoportal,
- publishing of links to existing Web pages only putting URL of Web pages to extractor,
- publishing directly new web pages stored in zip file. These web pages are directly accessible through the geoportal.

## 7. CONCLUSIONS

The tools of the proposed training platform are approved and ready for practical use in the educational process. The training platform includes social networking tools including virtual libraries, educational materials, content sharing tools, connection to social networks and application integrator. In the middle of this platform stands the BizBiz tool as a modern approach to the distance video e-learning. The advantage of this approach is the ease of use and setup, as well as low restrictions on hardware, because ordinary PC with a broadband connection can be used for conferences up to several tens of participants. The costs for the users are virtually irrelevant. The principles of training platform and described tools are broadly applicable in domain of e-learning.

## REFERENCES

[1] T. O'Reilly, "What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software".
[2] D. Osimo, "Web 2.0 in Government: Why and How?" JRC Scientific and Technical Report, European Communities, 2008, ISSN 1018-5593.
[3] A. Dzerve, R. Berzins, K. Janecka, K. Charvat, "SDI-EDU – Training Environment Implementation", Technical report, SDI-EDU for regional and urban planning project, 2010.
[4] K. Janecka, K. Charvat, "Analysis of Existing Tools and Content", Technical report, SDI-EDU for regional and urban planning project, 2010.
[5] K. Janecka, R. Berzins, K. Charvat, A. Dzerve, "On How to Build SDI Using Social Networking Principles in the Scope of Spatial Planning and Vocational Education", B. Murgante et al. (Eds.): ICCSA 2011, Part II, Lecture Notes in Computer Science 6783, pp. 78–92, 2011.

# Web Based Collaboration in Teaching Teamwork

Shan Barkataki
Computer Science, California State University
Northridge, CA 91330, USA

Robert Lingard
Computer Science, California State University
Northridge, CA 91330, USA

**ABSTRACT**

Teamwork is recognized as an important skill for engineering and computer science students. Industry strongly advocates inclusion of teamwork in the curriculum and teamwork is a requirement for ABET accreditation. Unfortunately most engineering and computer science programs fail to address teamwork as something that needs to be taught. It is usually expected that students should learn teamwork skills on their own, through participation in various team projects. Little thought or effort is given as to how to improve the way teaching is done in order to improve students' abilities to function on teams. One reason for this is that teaching teamwork demands a good deal of time and effort from the faculty. This is especially the case when teamwork is incorporated in regular lecture-discussion classes, rather than classes dealing with capstone projects. The result is that students often do not learn the critical teamwork skills such as: communication, planning and tracking, collaborative design, and peer reviews. We have developed an approach for teaching effective teamwork skills utilizing Yahoo Groups along with a Learning Management System (LMS), such as Moodle. Our approach improves student learning of teamwork skills by assuring that effective communication and team collaboration develops among all team members. Instructors are able to teach teamwork skills without having to devote a great deal of extra time and effort. Experience with this approach has shown more sustained participation by all team members in performing the collaborative work and greatly improved team communications.

**Keywords:** Agile Processes, Assessment, Collaboration, Communication, Teamwork

## 1. INTRODUCTION

Increasingly, teamwork skills are seen as vital for today's engineers. Michelle Shearer, National teacher of the year, recently addressed high school students attending a STEM (Science, Technology, Engineering and Math) seminar and identified what she called the four C's that are essential in STEM careers – critical thinking, creative problem solving, communication and collaboration. This view she feels "shatter[s] the image of what students think a traditional scientist or engineer looks like" [1]. Notice that two of the four C's cover communication and collaboration. As much as we tend to believe the importance of these last two C's, we do little to actually teach our students these skills. Ten years ago, Berry and Lingard reported that although "many programs today make team projects fundamental elements of their curricula, few actually teach teamwork and communication skills directly" [2]. As Purzer has more recently said, "Despite the importance of gaining effective teaming skills for our students and the increased popularity of using collaborative learning methods in college classrooms, there are limited modules and instructional tools designed to teach teaming skills to engineering students" [3]. The focus of our work is on developing effective techniques by which students can learn and improve teamwork skills and additionally provide the means by which instructors can evaluate and assess the degree to which individuals are learning these skills.

## 2. BACKGROUND

Previous efforts in attempting to assess student learning of teamwork skills using peer evaluations have provided some incite regarding which teamwork skills seem to be most difficult to learn [4]. After identifying a long list of important teamwork skills, the ones in which students rated their fellow teammates the poorest included the ability to clearly communicate with other team members and the abilities to both ask for help and to give help when needed. These are precisely the communication and collaboration skills deemed essential for today's engineers and scientists and are the skills targeted by our approach.

Two of the main challenges in teaching teamwork are dealing with negative student perceptions about teamwork and addressing the workload issues of faculty. Students have been conditioned to work independently and often view team assignments as extra work in which some students do little and yet receive credit based on the efforts of others. Often the advantages of teamwork are overshadowed by these negative aspects. Studies have shown that poor experiences on teams reinforces students' perceptions that teams are more trouble than they are worth [5].

The other challenge is that faculty view teaching teamwork as additional work, and furthermore, many feel unqualified to teach or evaluate teamwork skills. Many instructors have never worked in a team environment, and teaching teamwork skills does not come naturally. Certainly, training faculty can help to improve the teaching and evaluation of collaboration skills, but one goal of our approach is to make it easier for faculty to provide meaningful teamwork assignments and to facilitate the evaluation of both team and individual achievement.

## 3. APPROACH

We have developed an approach for teaching effective teamwork skills utilizing web collaboration. Within an agile framework [6] and using tools like Yahoo Groups we can help students to experience a successful teamwork activity. This

approach can be used to help students and faculty achieve the following goals:

- Engage all students in the development of a well defined and significant problem that is understood by all
- Help form cohesive teams
- Ensure development of appropriate plans and schedules
- Maintain effective communication among all team members
- Ensure that all team members are engaged in performing technical work and in the team decision making process, including issues related to quality control and configuration management
- Engage students in tracking execution of project activities with respect to the plans
- Optimize learning of teamwork skills by focusing on technical work products and peer reviews rather than production of formal documents and presentations
- Promote fair and accurate grading of individual students reflecting level of participation and the quality of the work products produced
- Avoid excessive workload for faculty in managing and grading the team projects

This approach improves student learning of teamwork skills by assuring that effective communication and team collaboration develops among all team members. Our approach leverages on using an Agile development process that encourages formation of self-organizing teams and effective team communication [7 ]. Most of the team interactions occur asynchronously, using web based tools such as Yahoo Groups, with short face-to-face meetings taking place once a week. The features available in web based collaboration tools like Yahoo Groups allow both the mentoring and monitoring of project teams. Faculty can easily assess the levels of participation by the individual members and also gauge team progress as the project develops. Our approach enables faculty to teach teamwork skills without demanding a great deal of extra time and effort.

## 4. CONCLUSIONS

We have applied this approach in teaching teamwork in a number of non-capstone classes. We have achieved good results, as evidenced by sustained participation by all team members in collaborative work and development of good team communications. Our presentation will describe this approach in greater detail and share the results we have obtained to date.

## 5. REFERENCES

[1] M. Neal, "STEM Seminar Creates Pathway for Students", FrederickNewsPost.com, May 25, 2011.

[2] E. Berry, R. Lingard, "Teaching Communication and Teamwork in Engineering and Computer Science", **Proc. 2001 ASEE Annual Conf.**, Albuquerque, NM, June 2001.

[3] S. Purzer, "The Merit Kit: Methods For Evaluating Roles And Interactions In Teams", **Proc. 2010 ASEE Annual Conf.**, Louisville, KY, 2010.

[4] R. Lingard, "Teaching and Assessing Teamwork in Engineering and Computer Science", **Proc. International Symposium on Engineering Education and Educational Technologies: EEET 2009** in the context of The 2nd International Multi-Conference on Engineering and Technological Innovation: IMETI, Orlando, FL, 2009.

[5] M. D. Aldridge, et al, "Introduction to team-based design for students in engineering, business, and industrial design", Final report to the National Science Foundation, 1996.

[6] K. Beck, et al, "Manifesto for Agile Software Development", http://agilemanifesto.org/, 2001.

[7] M. Cohn, **Succeeding with Agile: Software Development Using Scrum**, Addison-Wesley Professional, 2009.

# A Bachelor of Science in Modeling and Simulation Engineering

James F. Leathrum, Jr. and Roland R. Mielke
Department of Modeling, Simulation and Visualization Engineering
Old Dominion University
Norfolk, VA 23529, USA

## ABSTRACT

In recent years, a number of Modeling and Simulation graduate programs have been initiated to address the growing demand in industry and government for engineers and scientists educated in this discipline. However, much less has been done to develop undergraduate programs. This paper addresses the design of a curriculum for a baccalaureate program in Modeling and Simulation Engineering. The constraints and objectives that guide the curriculum design are described. Then the development of a new curriculum, designed to meet the ABET accreditation requirements for general engineering programs, is presented.

**Keywords:** Higher Education, Undergraduate Program, Modeling and Simulation, Engineering Program, Curriculum Design.

## 1. INTRODUCTION

An NSF Blue Ribbon Panel on Simulation-Based Engineering Science (SBES) [1] identified the importance of simulation in engineering and science research, but noted that the United States' leadership in the field is losing ground. The panel also identified the challenges involved in SBES, including the "education of the next generation of engineers and scientists in the theory and practice of SBES," and proposed a "sweeping overhaul of our educational system." Similar findings were presented by the National Research Council [2] where it was recommended that "DoD should ensure that its educational programs provide experiences in which students integrate the activities of modeling, simulation, analysis, evaluation, and communication…", going as far as to propose a core curriculum applicable to DoD practitioners. Congress has attempted to address these pressing needs by passing legislation [3] for a grant program to create new M&S degree programs. While no funding was appropriated, development of M&S academic programs was included in the Department of Education's Fund for the Improvement of Postsecondary Education (FIPSE) Comprehensive Program [4] proposal solicitation in 2010. However, no awards for M&S program development were awarded.

Since the late 1990's, a number of papers have been written stating the importance and urgency for developing educational programs in modeling and simulation. These papers identify desirable program outcomes [5], present suggestions for course and curriculum content [6], [7], and describe potential approaches for, and challenges in, implementing a modeling and simulation program [8], [9], [10], [11]. More recently, curricula [12], [13] and models [14] for graduate modeling and simulation programs have been described. Graduate modeling and simulation programs have been started at several universities including the University of Alabama – Huntsville [15], Arizona State University [16], California State University – Chico [17], Georgia Institute of Technology [18], Old Dominion University [19], and the University of Central Florida [20]. At the undergraduate level, several universities have developed tracks or concentrations focusing on narrow sub-areas of modeling and simulation as part of other degree programs. However, to date, no ABET accredited engineering program in modeling and simulation has been fully implemented

This paper describes the development of the first undergraduate degree program in modeling and simulation [19]. The program is being implemented, one program year each calendar year, over the period 2009 - 2013. The program will award the Bachelor of Science Degree in Modeling and Simulation Engineering (M&SE); it is anticipated that the first graduates will complete their studies in Spring 2013. The curriculum is designed to meet the ABET accreditation requirements for a general engineering program. Graduates will be prepared for employment as entrance level M&S engineers and for graduate study in modeling and simulation and possibly other closely related disciplines. Graduates also will be prepared to seek certification as a Certified Modeling and Simulation Professional (CSMP) and, with appropriate selection of curriculum electives, licensure as a Professional Engineer (PE).

The focus of the paper is on the design of the M&SE curriculum. The curriculum is designed to achieve four goals: (1) to satisfy the ABET criteria for a general engineering program; (2) to provide appropriate coverage of the M&S Body of Knowledge; (3) to achieve discipline-specific student outcomes; and (4) to meet university and college requirements for baccalaureate programs. In Section 2, the curriculum requirements imposed by these goals are described. Each goal is investigated separately and the impact on the curriculum design is noted. In Section 3, the details of the curriculum design process and the resulting M&SE curriculum are presented. Finally, in Section 4, the remaining activities needed to complete and document the curriculum course designs are summarized.

The purpose of the paper is two-fold. First, the description of an M&SE curriculum may serve as a useful model for other universities considering the initiation of similar programs. Second, it is hoped that the paper will promote additional discussion in the literature concerning the development of undergraduate modeling and simulation programs.

## 2. CURRICULUM REQUIREMENTS

There are at least four sources of curriculum constraints that should be addressed in the design of a modeling and simulation engineering curriculum. These sources include the ABET criteria for accrediting engineering programs, the literature defining an M&S body of knowledge, a set of discipline-specific student outcomes identified by program faculty, and university general education requirements. In this section, these requirements are described and the impact on curriculum structure is noted.

**ABET Criteria**

The M&SE program is an undergraduate engineering program designed to meet accreditation standards of the Engineering Accreditation Commission of ABET. ABET specifies eight general criteria for accreditation of baccalaureate programs [21]. Three of these criteria, Criteria 2, 3, and 5, have a direct impact on curricula design and are identified here. At the present time, ABET does not recognize modeling and simulation engineering as a distinct engineering discipline; therefore, the M&SE program will seek accreditation under the guidelines for general engineering programs. For general engineering programs, there are no additional program-specific criteria. It is anticipated that eventually modeling and simulation engineering programs will become recognized as a separate engineering discipline and at that time will be assigned program-specific criteria. The following ABET criteria directly impact the design of the M&SE curriculum.

**ABET Criterion 2 - Program Educational Objectives.** Program educational objectives are defined as statements that describe the accomplishments expected of graduates during the first few years following graduation. Clearly, the program curriculum must be designed to foster the attainment of the program objectives. The program objectives defined for the M&SE program are enumerated in the following.

The M&SE program seeks to prepare engineering graduates who:

1. Practice modeling and simulation engineering successfully in different professional settings;
2. Work responsibly and effectively as members of a professional team or organization; and
3. Demonstrate an ability to adapt to new methods and technologies, and to changing career challenges.

**ABET Criterion 3 - Student Outcomes.** Student outcomes are defined as statements that describe what students are expected to know and be able to do by the time of graduation. Once again, the program curriculum must be designed to continually advance students toward achieving the desired outcomes. The student outcomes defined for the M&SE program are enumerated in the following.

M&SE students who qualify for graduation will have:

1. An ability to apply knowledge of mathematics, science, and engineering;
2. An ability to design and conduct experiments, as well as to analyze and interpret data;
3. An ability to design a system, component, or process to meet desired needs within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability;
4. An ability to function on a multidisciplinary team;
5. An ability to identify, formulate, and solve engineering problems;
6. An understanding of professional and ethical responsibility;
7. An ability to communicate effectively;
8. The broad education necessary to understand the impact of engineering solutions in a global, economic, environmental, and societal context;
9. A recognition of the need for, and an ability to engage in life-long learning;
10. A knowledge of contemporary issues; and
11. An ability to use the techniques, skills, and modern engineering tools necessary for engineering practice.

**ABET Criterion 5 – Curriculum.** The curriculum criterion identifies broad content areas, and credit allocation for these areas, that must be included in an engineering curriculum. The engineering curriculum must include:

1. One year of a combination of college level mathematics and basic sciences, some with experimental experience, appropriate to the discipline;
2. One and one-half years of engineering topics, consisting of engineering sciences and engineering design appropriate to the student's field of study;
3. A general education component that complements the technical content of the curriculum and is consistent with the program and institution objectives; and
4. A major design experience based on the knowledge and skills acquired in the earlier course work and incorporating appropriate engineering standards and multiple realistic constraints.

**M&S Body of Knowledge**

A number of significant efforts have been made to define an M&S body of knowledge [6], [22]. However, the focus of these efforts primarily has been to identify content for development of M&S graduate programs or for development of licensure requirements for current M&S practitioners. Thus, while existing body of knowledge presentations may not be entirely appropriate to undergraduate program development, the work serves as a framework from which to select a subset of content areas that are appropriate to M&SE baccalaureate programs.

Presentation of the M&S body of knowledge often consists of a listing of content areas displayed in outline form. The body of knowledge listing displayed here is taken from [23] and was used as the basis for the 2009-2011 revision of the CMSP examination. The major content area labels are shown as headings in this listing while second level topic areas are presented in list form under the appropriate heading.

1. Fundamental Concepts and Context
   Fundamental terms and concepts, categories and paradigms, history, communities-of-practice, M&S related disciplines
2. Applications and Domains
   Applications, domains
3. Modeling Methods
   Physics based models, finite element models, human behavior models, Monte Carlo simulation, discrete event simulation, continuous simulation
4. Simulation Implementation
   M&S life cycle, modeling standards, development process, specialized languages, verification & validation, distributed simulation, multi-resolution

simulation, semi-automated forces, virtual environments, virtual reality

5. Supporting Tools, Techniques, and Resources
   Probability and statistics, mathematical methods, software engineering, simulation environments, model and simulation resource repositories, M&S organizations
6. Business and Management of M&S
   Ethics and principals for practitioners, management of projects and processes, workforce development, business practice and economics, industrial development

**Discipline-Specific Essential Knowledge and Skills**

The M&SE program should prepare engineering graduates who can utilize modeling and simulation in various domains and for different applications, and who possess the foundation upon which to expand the current M&S body of knowledge. The M&SE program faculty has defined a set of essential knowledge and skills that they believe form the technical foundation for the discipline of modeling and simulation engineering. These are the concepts, principles, and methods that anchor the M&SE curriculum; they represent the fundamentals that every M&SE graduate must know and be able to use. The M&S essential knowledge and skills are stated as a set of student outcomes that are focused on the technical components of the M&S curriculum. These student outcomes are stated in the following.

M&SE students who qualify for graduation will have:

1. An ability to communicate designs across technical and non-technical boundaries;
2. An ability to model a variety of systems from different domains;
3. An ability to develop an input data model based on observed data;
4. An ability to select and apply appropriate simulation techniques and tools;
5. An ability to develop simulations in software;
6. An ability to apply the experimental process to acquire desired simulation results;
7. An ability to apply visualization techniques to support the simulation process;
8. An ability to use appropriate techniques to verify and validate models and simulations; and
9 An ability to analyze simulation results to reach an appropriate conclusion.

**University Requirements**

The vast majority of universities offering baccalaureate programs specify general education requirements that must be met by all undergraduate curricula at that university. These requirements vary significantly, both in content specified and credits required, from one institution to another. In addition, colleges of engineering often place additional requirements or constraints on engineering program curricula.

The M&SE curriculum described here is designed to accommodate both university and college requirements. The general education requirements consist of approximately 56 credits of courses; some of these courses address fundamental skills needed throughout the program of study and must be scheduled early in the curriculum. Other general education courses address application of these fundamental skills and thus must be scheduled later in the curriculum. Fortunately, a significant subset of the general education content has overlap with the ABET requirements and, to a lesser degree, the M&S Body of Knowledge and the discipline-specific student outcomes. Only in this way is it possible to meet all requirements in a single curriculum consisting of reasonable number of credits. The college of engineering requires that all engineering curricula share a common freshman curriculum. The purpose of this requirement is that it affords engineering freshmen the flexibility to investigate several different engineering disciplines before having to declare a major. The general education requirements and the college requirements together place rather significant constraints on the design of the M&SE curriculum.

## 3. CURRICULUM DESIGN

The curriculum development presented here has not followed one particular curricular model, but rather has attempted to take advantage of the best features of several approaches. A balance has been struck between the use of an objectives model and a process model [24], [25]. The objectives model begins the design process with a statement of student outcomes and best fits the presentation of the ABET requirements and the discipline-specific requirements. However, the process model better addresses the desire to teach problem solving and technical innovation. The development process most closely followed the model of Barnett, Parry, and Coate [26], based on three domains: knowledge, action, and self, and that of Integrated Course Design (ICD) [27] based on learning goals, teaching and learning activities, and feedback and assessment. While the curriculum presented is new and the first of its kind to be put into practice, much thought has been placed on how curriculum outcomes, the M&S Body of Knowledge, and assessment define the curriculum.

The M&SE curriculum is first and foremost an engineering program having a focus on problem solving, design, and experimentation. The curriculum is designed with 128 credits; 32 credits of mathematics and basic science courses, 58 credits of engineering science and design courses, 32 credits of general education courses, and 6 credits of approved electives. The credit distribution is selected to satisfy ABET Criterion 5 and the university's general education requirements. The curriculum is displayed in "showcase" format in Figure 1. In this display, the courses are distributed over eight semesters and the courses are sequenced to satisfy all prerequisite and co-requisite requirements. The core technical courses are displayed in Figure 2. In this display, the core courses are grouped into three content areas: modeling and simulation; software development; and analysis. This display also shows explicitly all prerequisite and co-requisite relationships among the courses.

The following statements identify important characteristics of the M&SE curriculum that were considered carefully as during curriculum design.

- The basic skill courses, along with the mathematics and basic science courses, are scheduled early in the curriculum. This is done to address the ABET requirement that fundamental math and science be introduced early and then utilized throughout the curriculum. It also facilitates the requirement for a common freshman engineering curriculum and the general education requirement that basic skills courses be completed in the freshman and sophomore years.

- The core engineering science and engineering design courses are grouped into three content areas: modeling and simulation; software development; and analysis. The modeling and simulation content area consists of courses that address the design and implementation of models and simulations. This track includes courses on: Monte Carlo simulation; discrete event simulation; continuous simulation; and modeling methodologies. The software development content area consists of courses that develop a programming capability and then use that capability to code simulations. Background courses are computer science courses while simulation development is an M&SE course. Computer visualization is included in this content area. The analysis content area consists of the courses that lead to an analysis capability. This area includes the mathematics courses in calculus, differential equations, and probability and statistics. It also includes a simulation analysis course that addresses random number generation, input data modeling, output data analysis, verification and validation, and experimental design. These core areas are designed to address the ABET technical outcomes, the discipline-specific outcomes, and a large component of the M&S Body of Knowledge.
- A skills content area consists of the courses that provide essential academic skills such as the ability to communicate effectively. This content area addresses the requirements of the ABET outcomes, the general education requirements, and to a lesser degree the M&S Body of Knowledge and the discipline-specific outcomes. The communication skills component is particularly strong and consists of the following course sequence: English composition, technical writing, public speaking, and information literacy and research. Extension and practice of these skills also are present in a number of the core technical courses.
- A professional skills and knowledge content area addresses the ABET requirements and the M&S Body of Knowledge components on professional conduct and business processes. This content area includes the following courses and topics: introduction to engineering profession; engineering ethics; impact of technology; project management; fundamentals of engineering examination preparation; and the capstone design experience.
- The curriculum design allows for inclusion of a formal minor; that is, a set of 12 credits at the upper division offered and approved by an external academic department. We are working with several departments to organize minors that will have a beneficial impact on the student's major studies. For example, the Psychology Department has a minor in human factors that includes courses on perception, cognition, human-computer interfacing, and human behavior modeling. The Biomedical Engineering minor includes courses in biomechanics, human physiology, and innovation and entrepreneurship in the biomedical area. For those students wishing to seek licensure as a professional engineer, minors in civil, electrical, computer, and mechanical engineering are available. The availability of a minor in the M&SE curriculum is viewed as a strength and students are encouraged strongly to utilize this option.

**Figure 1. Showcase M&SE Curriculum.**

| Course Number | Course Title | Credits |
|---|---|---|
| | Freshman Year – First Semester | |
| MATH 211 | Calculus I | 4 |
| ENGL 110C | English Composition I | 3 |
| CHEM 121N | Chemistry I & Lab | 4 |
| ENGN 110 | Engineering & Technology I | 2 |
| Gen Ed | Oral Communication | 3 |
| | Freshman Year – Second Semester | |
| MATH 212 | Calculus II | 4 |
| CHEM 123N | Chemistry II | 3 |
| CS 150 | Programming I | 4 |
| PHYS 231N | University Physics I | 4 |
| MSIM 111 | Information Literacy for M&SE | 2 |
| | Sophomore Year – First Semester | |
| STAT 330 | Probability and Statistics | 3 |
| PHYS 232N | University Physics II | 4 |
| CS 250 | Programming II | 4 |
| CS 252 | Introduction to UNIX | 1 |
| MSIM 201 | Introduction to M&S | 3 |
| | Sophomore Year – Second Semester | |
| MATH 307 | Differential Equations | 3 |
| ENGL 231C | Technical Writing | 3 |
| Gen Ed | Human Creativity | 3 |
| Gen Ed | Literature | 3 |
| MSIM 205 | Discrete Event Simulation | 3 |
| MSIM 281 | M&S Laboratory 1 | 1 |
| | Junior Year – First Semester | |
| CS 330 | Object-Oriented Prog. & Design | 3 |
| CS 381 | Discrete Structures | 3 |
| MSIM 320 | Continuous Simulation | 3 |
| MSIM 382 | M&S Laboratory 2 | 1 |
| Gen Ed | Human Behavior | 3 |
| Gen Ed | Option D Course 1 | 3 |
| | Junior Year – Second Semester | |
| MSIM 331 | Simulation Software Design | 3 |
| MSIM 383 | M&S Laboratory 3 | 1 |
| MSIM 351 | Analysis for M&S | 3 |
| MSIM 310 | System Modeling | 3 |
| Gen Ed | Interpreting the Past | 3 |
| Elect | Approved Elective 1 | 3 |
| | Senior Year – First Semester | |
| MSIM 441 | Computer Graphics & Visualization | 3 |
| MSIM 487 | Capstone Design I | 4 |
| Gen Ed | Option D Course 2 | 3 |
| ENMA 401 | Project Management | 3 |
| MSIM 4xx | Approved MSIM Elective 1 | 3 |
| | Senior Year – Second Semester | |
| ENMA 480 | Engineering Ethics | 3 |
| MSIM 488 | Capstone Design II | 3 |
| Elect | Approved Elective 2 | 3 |
| MSIM 4yy | Approved MSIM Elective 2 | 3 |
| ENGN 401 | FE Exam Review | 1 |
| Gen Ed | Impact of Technology | 3 |

## 4. CONCLUSION

A brief overview of the development of a curriculum for an undergraduate program in Modeling and Simulation Engineering is presented. The constraints that the curriculum must satisfy are described. The resulting curriculum is displayed and the more significant features of the curriculum are identified. The next steps in the design process include the design of course content, development of curriculum content maps, mapping the content back to the curriculum requirements, and developing appropriate assessment strategies for measuring the degree of success in achieving the student outcomes and objectives. These next steps are in progress.

## REFERENCES


[1] "Simulation-Based Engineering Science," Report of the National Science Foundation Blue Ribbon Panel on Simulation-Based Engineering Science, National Science Foundation, May 2006.

[2] National Research Council, "Defense Modeling, Simulation, and Analysis: Meeting the Challenge," The National Academies Press, Washington, DC, 2006.

[3] Scott, R. C., "HR 4165 - To provide grants to encourage and enhance the study of modeling and simulation at institutions of higher education," The Congressional Record, November 13, 2007.

[4] Department of Education, "Overview: Information Fund for the Improvement of Postsecondary Education (FIPSE) - Comprehensive Program," *Federal Register*, June 14, 2010.

[5] Rogers, R., "What Makes a Modeling and Simulation Professional: The Consensus View from One Workshop," Proceedings of the Winter Simulation Conference, Piscataway, NJ, 1997.

[6] Oren, T.I., "Toward the Body of Knowledge of Modeling and Simulation", Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC), Paper 2025, Orlando, FL, December 2005.

[7] Petty, M.D., "Graduate Modeling and Simulation Overview Course," Final Technical Report, Contract N00014-031-0948, Defense Modeling and Simulation Office, March 2006.

[8] Sarhoughian, H.S. and Zeigler, B.P., "Towards Making Modeling and Simulation into a Discipline,' 2000, web source: www.acims.arizona.edu.

[9] Nance, R.E., "Simulation Education: Past Reflections and Future Directions", Proceedings of the 2000 Winter Simulation Conference, pp. 1595-1601, 2000.

[10] Nance, R.E. and Balci, O., "Thoughts and Musings on Simulation Education", Proceedings of the Winter Simulation Conference, pp. 1567-1570, 2001.

[11] Roberts, C.A. and Ghosh, S., "A Proposed Model for an Undergraduate Engineering Program in Modeling and Simulation", Proceedings of the International Conference on Simulation Education, Western Simulation Multi-conference, pp. 38-44, 2004.

[12] Szczerbicka, H., Banks, J., Rogers, R., Oren, T., & Sarjoughian, H.Z., "Conceptions of Curriculum for Simulation Education," 2000 Winter Simulation Conference, pp. 1635-1644, 2000.

[13] Sarhoughian, H., Cochran, J., Collofello, J., Goss, J., and Zeigler, B. "Graduate Education in Modeling & Simulation: Rationale and Organization of an Online Masters Program," Proceedings of the Summer Computer Simulation Conference, 2004.

[14] Mielke, R., Scerbo, M., Gaubatz, K., and Watson, G., "A Model for Multidisciplinary Graduate Education in Modeling and Simulation,". International Journal of Simulation and Process Modeling, Vol. 5, No. 1, pp. 3-13, February 2009.

[15] University of Alabama – Huntsville, web source: http://catalog.uah.edu, accessed May 25, 2011.

[16]Arizona State University, web source: http://www.asu.engineeringonling.com, accessed May 25, 2011.

[17] California State University – Chico, web source: http://www.ecst.csuchico.edu/~mcleod/degree/html, accessed May 25, 2011.

[18] Georgia Institute of Technology, web source: http://www.msrec.gatech.edu, accessed May 25, 2011.

[19] Old Dominion University, web source: http://eng.odu.edu/msve, accessed May 25, 2011.

[20] University of Central Florida, web source: http://www.ist.ucf.edu/phd, accessed May 25, 2011.

[21] ABET Engineering Accreditation Commission, "Criteria for Accrediting Engineering Programs", 2011-2012 Accreditation Cycle, web source: http://www.abet.org, accessed May 23, 2011.

[22] Birta, L.G., "The Quest for the Modelling and Simulation Body of Knowledge," Proceedings of the Sixth Conference on Computer Simulation and Industry Applications, Tijuana, Mexico, 2003.

[23] Petty, M., "Modeling and Simulation Body of Knowledge Topics", in Certified Modeling & Simulation Professional Program – Management Plan, Modeling and Simulation Professional Certification Commission, Version 1.0, June 12, 2009, from http://www.simprofessional.org.

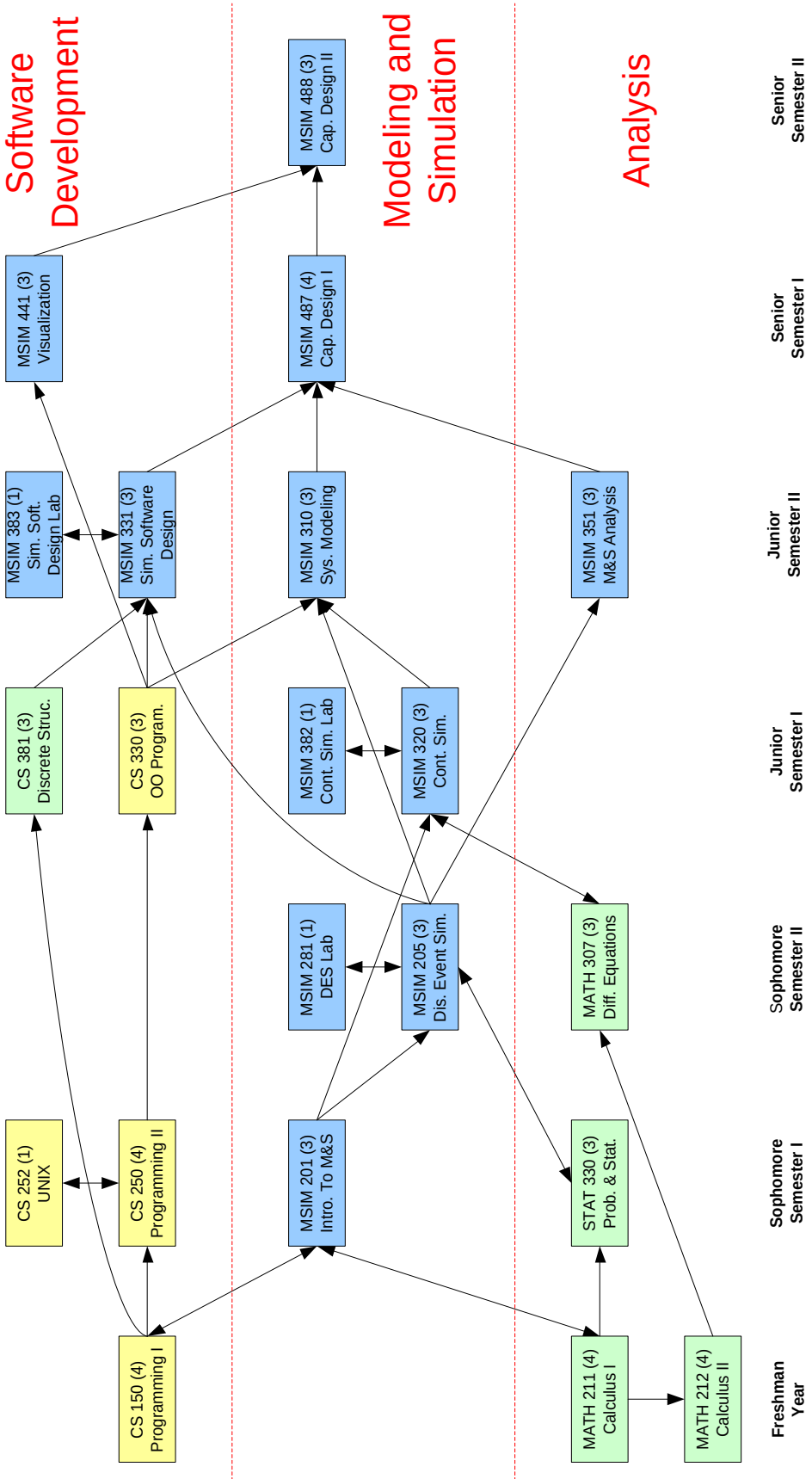
[24] Howard, J., "Curriculum Development," 2007, web source: http:// org.elon.edu/ catl/ documents/ Curriculum Development.pdf, Retrieved September 8, 2010.

[25] McKimm, J., "Curriculum Design and Development," 2007, web source: http:// www. faculty. londondeanery.ac.uk/ e-learning/ setting-learning-objectives/ Curriculum_design_and_development.pdf, Retrieved September 8, 2010.

[26] Barnett, R., Parry, G., and Coate, K., "Conceptualising Curriculum Change," Teaching in Higher Education , *6* (4), pp. 435-449, 2001.

[27] Fink, D., *Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses*, San Francisco: Jossey-Bass, 2003.

**Figure 2. Core M&SE Courses Showing Prerequisite Relationships.**



# Modeling & Simulation as an Enabling Technology for Engineering Innovation Education

James F. Leathrum, Jr. and Roland R. Mielke
Department of Modeling, Simulation and Visualization Engineering
Old Dominion University
Norfolk, VA 23529, USA

## ABSTRACT

The inclusion of innovation within engineering education is being identified as a crucial effort to ensure our technological competitiveness as evident by the number of solicitations including engineering innovation education by the National Science Foundation. While most work is spent on identifying how to create environments that foster innovation, this paper focuses on developing a technical skill set that enables innovation when in the correct environment. The premise of this paper is that the skills taught within the modeling and simulation curricula are those that enable innovative thinking. The skill set focuses on the abstract thinking, communication, and evaluation skills. This paper does not propose changing current engineering curricula, but rather to augment them by introducing abstract modeling concepts early in the curriculum, allowing students to see the generation of an innovative idea. This is be done through a series of small modules and assignments spread throughout the curriculum. Furthermore, the modules benefit the existing curricula by providing a context for why the students are learning the classic material.

**Keywords:** Innovation, Modeling and Simulation, Engineering Education

## 1. INTRODUCTION

The basis for this paper is our recognition that the concepts and methods that comprise the discipline of modeling and simulation are precisely the tools and skills required to conduct the process of innovation. Modeling skills are required to capture new ideas and systems, and then to communicate those ideas to other groups that often have vastly different backgrounds and points-of-view. Visualization is needed to help build mental models and understanding of how complex systems work and are organized. Simulation provides a cost-effective means to determine system behavior and performance and to investigate life-cycle issues important to a successful system design. It is our position that making an engineering student into an innovator will be enhanced or even enabled if that student has the tools needed to conduct the process of innovation. We are proposing a path for the introduction of M&S skills with an emphasis on enabling innovative ideas within existing STEM curricula.

Innovation is the complex process of introducing novel ideas into use or practice and includes entrepreneurship as an integral part [1]. Entrepreneurship involves the commitment of finance and business acumen in an effort the transform the new concepts of inventions into economic goods. The most obvious form of entrepreneurship is that of starting new businesses. While much is being done to encourage entrepreneurship, our focus is on the complex process of innovation, which is less understood yet is the most vital to comprehend.

This paper proposes developing new and novel approaches for teaching innovation in the STEM areas, especially engineering, that can be integrated into existing curricula with minimal effort. Should the introduction of innovation be a heavy burden on existing curricula with little immediate rewards, it is unlikely the proposed work will be adopted. We also believe that innovation and the associated skills need an early introduction for students to embrace them and follow educational paths leading to innovation. As a first step in the innovation education process, we identify M&S as the enabling technology allowing the reasoning, visualization, evaluation, and communication required in innovation. We propose

core skills from M&S that should be taught to all engineering students to improve the likelihood of fostering innovation.

The proposed engineering educational evolution follows the maturation process of M&S as a discipline in its own right. This is similar to the development of computer science as a discipline through the 1960's and early 1970's and the recognition that the skill set was required by engineers as well. As a result, a core competence in computer science was introduced into engineering curricula and became a requirement for accreditation. Likewise, M&S has matured over the last decade with the creation of graduate programs (Old Dominion University, University of Central Florida, University of Alabama - Huntsville, CSU - Chico, etc.), followed by the first undergraduate program being introduced in 2010 at Old Dominion University. Thus it follows that the time has come to evaluate the discipline's role in engineering education. This paper considers the introduction of M&S as a core competence to enable innovation across all engineering disciplines.

## 2. INNOVATION IN ENGINEERING EDUCATION

Some work has been done attempting to introduce an innovative environment in the engineering curriculum. [2-7] However, these primarily focus on developing an atmosphere and that fosters the opportunity to be creative, but ignore the engineering skill set that best enable engineers to be innovative in a timely and cost efficient manner. Steiner does attempt to identify the skills that make people innovative, but through querying managers, she really concludes on personal skills (energetic, enthusiastic, competitive, etc.) not technical skills (potentially including abstract modeling, top-down design, rapid prototyping, etc.).

Other works attempt to define the innovative process. [8-12] But again they focus on creating an atmosphere that fosters creativity or identifying what techniques have worked. In attempting to identify the characteristics of teams that have been successful in innovations, Petre identified that the teams were small, multidisciplinary, connected to academic institutions, intense workers, and visionary. Again, these are characteristics, not identifying appropriate skill sets in which engineers should be educated.

This paper attempts to identify a skill set to enable innovative thinking in engineers when placed in an appropriate environment. The skill set focuses on the abstract thinking, communication, and evaluation skills. Current engineering education is primarily based on analysis, teaching design in a bottom-up fashion by introducing low level theory and building up to larger designs of systems that already exist. This paper does not propose changing that basic philosophy, but rather to augment it by introducing abstract modeling concepts early in the curriculum, allowing students to see the generation of an innovative idea. This can be done through a series of small modules and assignments spread throughout the curriculum. The modules would work in the opposite direction from the classic bottom-up method, but would augment the existing curricula by providing a context for why the students are learning the classic material.

## 3. M&S CONCEPTS IN INNOVATION ENGINEERING

Problem understanding is important in recognizing the intricate and nuanced issues involved with identified problems. Once understanding is achieved, problem description requires unambiguously documenting the problem so that it may be easily communicated. Then, recognizing that the problem may be defined within certain known categories of problems will help support finding a solution to the problem or problem resolution. Properly investigating these creative processes and facilitating them through application of proven M&S tools and techniques fills an existing gap in the research of innovation.

We propose the following concepts as the core competencies in innovative engineering:

- *needs assessment* - the act of capturing the needs of a problem
- *conceptualization* - the creative act of formulating an idea.

- *visualization* - the act of creating a representation of an idea for better understanding.
- *evaluation* - the testing of an idea, may include engineering, manufacturing, and marketing feasibility.
- *communication* - the act of presenting the idea to others, potentially for collaboration or marketing.

While methodologies for teaching these competencies is not addressed in this paper, the importance of M&S to enable each is elaborated.

**Needs Assessment**

Needs assessment involves the definition of the requirements of a problem for which innovative ideas are sought. In engineering, this includes development of the problem specification. Frequently this involves documentation of ideas in writing. M&S attempts to formalize this process by capturing specifications as models. The needs are captured as behavioral models. Developing executable simulations of these models then allow the engineer to study if the proposed behavior will truly meet the needs. The model then acts as the specification that an innovative concept must meet, and provides a basis to validate conceptualized ideas against system needs.

**Conceptualization**

In conceptualization, it is important to be able to capture and manipulate one's ideas. The easier that this is done, the more opportunity for innovation. As in art, the ease of manipulating a medium facilitates the creative process. When developing a sculpture in stone, it is best to develop prototypes in an easier medium - drawings, clay, etc. - allowing the creative process to rapidly refine ideas.

M&S offers similar opportunities in the development of innovative ideas. M&S provides a fast, cheap, and safe development environment. An engineer can rapidly study alternative ideas. Unfortunately, most engineers only learn design tools applicable to their problem domain, and not general M&S techniques. This limits the opportunities for innovation to the capabilities of the tools at their disposal. Through developing general M&S skills, engineers could consider a broader realm of ideas. M&S techniques allow studying concepts at a behavioral level prior to design.

Modeling provides a means for engineers to capture their ideas. The simple act of model development often promotes a better understanding of a system. It also enables the definition of interfaces between subsystems within a system model, allowing a partitioning in the idea development. Then by simulating the system, a deeper understanding of system behavior is possible.

**Visualization**

The ability to visualize an innovative idea is assists in idea development and is crucial to communicating the idea to potential investors. M&S utilizes visualization heavily in understanding system behavior. Physical models support providing a visual understanding of a design. Data visualization and animation provide the ability to study system behavior as a function of time.

**Evaluation**

Evaluation is a prime emphasis of simulation. Once an idea is captured by a model, a simulation creates an executable environment in which the behavior of the modeled system can be studied. This allows rapid prototyping of ideas necessary to ensure its functionality. This is crucial when looking for venture capital or pursuing a patent. The utility and functionality of the idea must be clearly demonstrated. Simulation provides a safe, cost effective approach to achieving these goals.

However, frequently engineers do not fully understand simulation tools they employ. They often lack knowledge of simulation characteristics such as fidelity and time management. Often when a model is developed at an inappropriate level of fidelity, desired simulation features are missed. When dealing with time-based simulations, it is often necessary to know whether a simulation is continuous or discrete-event to fully understand results. For continuous simulations it is also necessary to know the unit of time advancement since system changes that occur in a shorter unit of time may not be observed. For discrete-event simulations, it is necessary to understand what constitutes and event to understand when state changes can occur.

**Communication**

The ability to communicate innovative ideas is crucial to their success. Communication is considered at two levels, the model level and the system behavior level. A model is a language for communication within M&S, enabling engineers to communicate ideas in the development stages, refining those ideas. Models provide a more formalized definition than written descriptions, largely removing the inherent ambiguities found in written documentation. Models can include mathematical models, state transition models, and even hybrid models. Engineers versed in general modeling techniques can then easily communicate ideas. Then when the models are executed as simulations, system behavior is captured and available to communicate requirements and/or capabilities. Thus visualization becomes a form of communication, capturing time varying behavior so that others, including non-engineers such as investors, can readily understand system behavior based on the idea.

## 4. IMPLEMENTATION CONCEPT

It is proposed that educational content for the teaching of innovation in the STEM areas, especially engineering, be designed so that it can be integrated into existing curricula with minimal effort. Should the introduction of innovation be a heavy burden on existing curricula with little immediate rewards, it is unlikely the educational effort will be adopted. We also believe that innovation and the associated skills need an early introduction for students to embrace them and follow educational paths leading to innovation. Learning modules will be developed that can be used to teach these tools and skills for different disciplines at various educational levels beginning at the high school level and continuing through the undergraduate college level. The focus is to introduce M&S concepts in the context of innovative ideas tailored to reinforce what the student is already learning.

Modules for innovation and entrepreneurship will be developed in self contained learning units so that they may be easily incorporated into existing courses or be combined to create whole courses that can be used as electives or that become part of a minor program. Furthermore, these modules will be adapted to fit the level of the learner at the high school, community college, and university. For instance, examples and tools will be characterized sufficiently so that collaborators at each level can choose appropriate instantiations for their level of instruction. Even within the college level, these modules may be adapted to fit specific disciplines such as the sciences and mechanical, civil, and aerospace engineering, etc. The results of efforts at this stage will be sets of M&S-enabled and engineering-oriented modules for both innovation and entrepreneurship.

The authors are working on the conceptualization of multidisciplinary innovation modules where the same problem domain can be utilized in all engineering disciplines. The modules can focus on relevant social problems such as homeland security, biomedical engineering, and alternative energy sources. The innovation modules will introduces through an engineering curriculum as single lectures and exercises in various courses, but with a single innovative concept strand spanning multiple courses and concepts. This allows the complete development of an innovative idea to be explored as a student's abilities are developed. However, while classical engineering is taught in a bottom-up strategy, the innovative concepts will be introduced as a top-down strategy. This has the added benefit of supplementing the standard engineering education process by providing a context for the content early in the curriculum.

An initial exercise develops the needs assessment, culminating in capturing a specification as a behavioral model. Ensuing exercises focus on idea conceptualization, evaluation, and prototyping. Modules will be linked between different engineering curricula so that the work done in one discipline is passed on to students in another discipline. This could possibly be carried into a senior design project, or lead into an elective course on the entrepreneurship process to provide a complete experience for the student.

## 5. CONCLUSIONS

While this effort is a work in progress, it is obvious that existing engineering curricula do not develop a skill set that promotes the out-of-the-box thought process necessary in the development of innovative ideas. However, the skill set produced by a curriculum in M&S does provide an appropriate skill set. Therefore, the introduction of M&S high-level modeling skills to develop behavioral models that are later refined to prototypes in the form of simulations, enables engineering students to partake in the innovation and entrepreneurship process. Without introducing these skills, the innovation process is more time consuming, costly, and risky.

## 6. REFERENCES


[1] *Invention: Enhancing inventiveness for quality of life, competitiveness, and sustainability*, Report of the Committee for Study of Invention, sponsored by the Lemelson-MIT Program and the NSF. Released April 23, 2004.

[2] Cropley, D., S. Priest and A. Cropley, "Fostering Creativity and Innovation in Engineering Students," *Putting Students Back into Engineering Education*, 1997.

[3] Li, W., J. Tsai, S. Tseng and I. Young, "Generation of New Ideas for Product Functions and Engineering Innovation Practices: A Hands-on Project Model for Machatronics Education," *Exploring Innovation in Education and Research*, Taiwan, March 2005.

[4] McKeag, D., "Embedding Creativity and Innovation in the Engineering Curriculum", Proceedings of the SEFI 2008 Annual Conference, July 2008.

[5] Perl, M., "Developing Creativity and Innovation in Engineering and Science," Proceedings of the Conference in Honor of C N Yang's 85th Birthday - Statistical Physics, High Energy, Condensed Matter and Mathematical Physics, Singapore, Nov. 2007.

[6] Steiner, C., "Educating for Innovation and Management: The Engineering Educators' Dilemma," *IEEE Transactions on Education*, vol. 41, issue 1, Feb. 1998, pp. 1-7.

[7] Yu, Y. and D. Shaw, "Fostering Creativity and Innovation in Engineering Students," 2006 International Mechanical Engineering Education Conference, Beijing, China, March, 2006.

[8] Petre, M., "How Expert Engineering Teams Use Disciplines of Innovation, " *Design Studies*, vol. 25, issue 5, Sept. 2004, pp. 477-493.

[9] Salter, A. and D. Gann, "Sources of Innovation in Engineering Design," *Research Policy*, vol. 32, issue 8, Sept. 2003, pp. 1309-1324.

[10] Yusuf, S., "From Creativity to Innovation," *Technology in Society*, vol. 31, issue 1, Feb. 2009, pp. 1-8.

[11] Cooper, J. (1998), "A Multidimensional Approach to the Adoption of Innovation," *Management Decision*, Vol. 36, Iss. 8, pp. 493-502.

[12] Garcia, R. and Calantone, R. (2002), A critical look at technological innovation typology and innovativeness terminology: a literature review. Journal of Product Innovation Management, 19: 110–132.

# ZAP!, Student Electricity Experiments Done in their Rooms

**Jerome PINE**
**Physics Dept, Caltech**
**Pasadena, California 91125, U.S.A.**

## ABSTRACT

An electricity laboratory course has been created at Caltech that is based on students doing experiments in their rooms, using parts and tools from a kit they receive at the start, for work during 20 weeks of the school year. The first laboratories are at a very basic level and the complexity increases over time. In the final experiment the students make a microwave transmitter and receiver and study radiation and detection by dipole antennas. During the course the physics of all the Maxwell equations is studied as well as circuit design and basic transistor electronics. Most of the students have little or no electronics knowledge at the start and they exhibit solid long-lasting learning.

**Keywords:** education, circuits, electricity, electronics, hands-on

## INTRODUCTION

Several years ago, John King and Phil and Phyllis Morrison at MIT developed an electricity course for freshmen called ZAP! It included laboratory experiments to be done by students in their rooms, elegantly invented by John King, and able to be done with relatively simple materials. Jerry Pine at Caltech built on the MIT course, made some adaptations and additions, and made it part of a first year elementary physics course taken by engineering and science students. The Caltech version described here is taught for twenty weeks in parallel with a standard electricity textbook course. There is a student book, about 110 pages long, to guide them through one experiment each week, intended to require on average about four hours of work. The experiment subject matter parallels the textbook, but the sequence of chapters studied is modified to reach electric circuits sooner. The experiments begin very simply and grow in complexity and sophistication during the course. The introduction to the student book lists four goals, summarized as follows:

*1. Abstract electrical concepts will be made real.*

*2. You will meet and become familiar with electrical technology.*

*3. You will get experience in solving real world "messy" experimental problems.*

*4. You will become designers, inventors, and doers.*

## THE AMBIANCE

The course begins with a distribution of kits to the students, which contain all the tools and parts needed for the 20 weeks of work, in a cardboard box about 30 x 40 x 16 cm. Figure 1 shows the contents laid out. The students buy the kit at approximately the cost of a textbook and after the course own a repertoire of tools, clip leads, and a multimeter for their ongoing use. An analog meter is used purposely, to provide experience with reading and interpolating from scales, and because it is an electromechanical application of physics that they study, and because it has circuitry they can understand. There is a complete set of tools and a soldering iron. The small parts, such as resistors, capacitors, LEDs, ICs, nuts and bolts, razor blades, etc. are in a plastic box. Power is supplied from a 12 volt a.c. plug-in transformer, an important feature for eliminating any shock hazard, and circuits are built on a

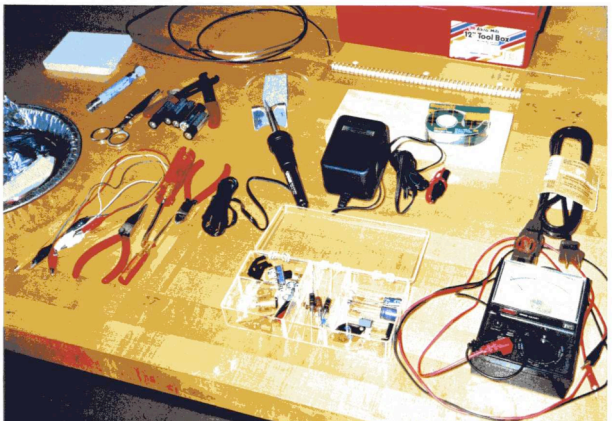


**Figure 1. The kit.**

perforated circuit board about 12 cm square. Other hardware, such as a sheet of styrofoam, a plywood baseboard, a pie pan, and so on, are also included. Connections are usually soldered, and the use of clip leads is minimized. A lab book, of 111 pages, leads the students through the experiments, but not in full detail. Component layout, measurement strategies, and trouble-shooting are left to them. They are encouraged to work in pairs, next to each other, each doing his or her experiment, but communicating to help each other and discuss what they are seeing. The lab book contains brief expositions of background theory when appropriate. Support for the students' experimentation comes from three sources in addition to the lab book:

1) They work independently but next to a partner, for discussion and for occasional help from extra hands.

2) There is a "help lab" each week in the evening before the student notebooks are collected for grading, staffed by faculty and grad students. Circuits that don't work are diagnosed and trouble-shot.

3) There is class time each week for preview comments and discussion of experimental results. In addition, since Caltech is a residential campus, there are upperclass students near at hand who have done the experiments and can provide advice. About half the freshmen do the labs, and the other half are in a different, more mathematical, course. Figure 2 below shows a student group at work in a dormitory room. There are three partners, which is not the usual rule.

**Figure 2. Students at work**

## THE EXPERIMENTS

There are thirteen experiments, listed below:

Simple Circuit Elements
Advanced Circuit Elements
Diodes and Rectifiers
Variable Voltage Power Supply
Power Supply Performance
The Thermistor and the Bridge
The Capacitor
The Transistor Amplifier
A Current Balance
Electromagnetic Induction
High Voltage Supply, the Arc
Microwaves
Digital Logic

Here are the highlights: The first three get the students used to using the meter and they study the properties of ohmic and non-ohmic circuit elements (resistors, diodes and LEDs). Then they make a variable voltage power supply with an IC regulator and determine how well it regulates. They use a bridge to calibrate the temperature dependence of the resistance of a thermistor and then determine its heat capacity for the next experiment. They measure the stored energy in a capacitor by discharging it through the thermistor and measuring the temperature rise. The transistor amplifier experiment is a quiz to be done alone, to evaluate how well they can design a circuit and use it to determine the current gain of a junction transistor. They do well, and they almost all succeed in accurately determining the current gain. That is the end of the first half of the course.

The remaining experiments are allotted two weeks each to complete, and are more technically demanding. The current balance is the classic measure of the force between parallel currents done with two parallel multi-turn coils and a very sensitive balance they make out of simple parts. The picture below shows a schematic view of it. The torque generated by nuts on the left end of the support bar balances gravity plus the force between the coils. This adds a bit of mechanical engineering

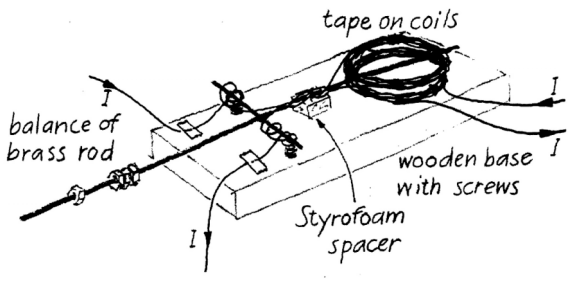
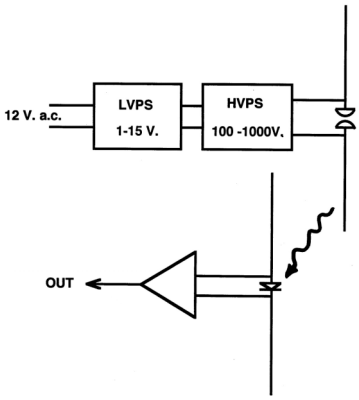



**Fig. 3 A current balance.**

to the work. They also are asked to analyze the net expected uncertainty in their force measurement, combined from multiple sources, and compare with the theoretical prediction. The induction experiment measures the a.c. coupling between the two coils used in the preceding experiment, and uses an opamp to amplify the induced voltage. They study the dependence on their spacing and the effect of an iron core in the center, and they need to make estimates of their expected results without their being a precise formula. They then build a high voltage power supply based on a 250 kHz r.f. oscillator, which produces more than 1000 volts at a safe high impedance. They study an arc between two closely spaced electrodes, and then use it to generate microwaves. The arc is coupled to a dipole antenna and the radiation is received by a matched antenna and detected with the help of a diode, opamp and meter. The sketch below shows this investigation diagrammatically, and following it is a picture of the apparatus, minus the pie plate reflector used behind the transmitting antenna.



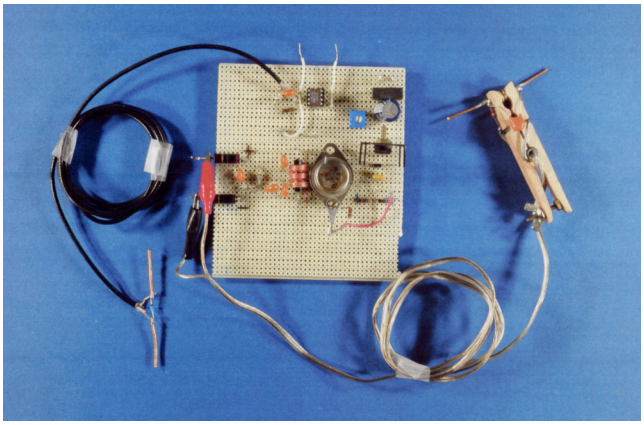



**Fig. 4 The microwave experiment**

Referring to the picture, many of the products of the twenty weeks of work are on the small circuit board. The variable d.c. power supply is in the upper right corner, and with a.c. input and fuse not shown. The big doorknob transistor powers the r.f. oscillator and its transformer, to the left, uses a few turns around a 5 mH inductor as the primary and the inductor windings as the secondary. There is a filter to generate d.c. from a voltage-doubler rectifier circuit. The high voltage is connected to two tacks between the jaws of a clothespin, which provides a micrometer adjustment of the transmitter arc length. The spark is pulsating because it is driven by an RC relaxation oscillation. The long black coax cable allows the receiving dipole to be moved around the transmitter to study the radiated field. A high frequency diode is across its arms. The rectified signal is amplified by an opamp circuit at the top center and read out on the multimeter. The students study the radiation pattern of the transmitting dipole in three dimensions. A final quiz experiment to be done alone follows, in which the students study CMOS transistor logic circuits. Like the other quiz, it is a new piece of learning and a test of their ability to design an experiment and interpret the results.

## CONCLUSION

Is experimental experience needed for learning about electricity? Or anything else, for that matter. Clearly, the answer is yes. Is ZAP! better than conventional "canned" labs. Clearly, our assessment quizzes would be far beyond the abilities of the students from that sort of experience. But do our students learn more electrical concepts better? We have no controlled experiments. A truly controlled experiment is not a practical option. Like most higher education innovations, faculty opinion is what we have. Five successive Caltech faculty members have volunteered to run ZAP, over 15 years. It is not a simple task, so they believe in it. Others in engineering also do. Our students continue in later years on a campus rich in opportunities for work in labs, and they will obviously be better at the arts of trouble shooting, improvisation, and working with a partner. Though some students balk at the start, most finish ZAP! with satisfaction and a feeling of accomplishment. The large majority say they would recommend it to another student.

ZAP! is a unique combination of ambiance and curriculum, to our knowledge only replicated in one instance, at École des Mines in Nantes, where it has become a key part of the curriculum for all students. The faculty has adapted it to match their goals and their curriculum and they are enthusiastic about it. Like Caltech, the Nantes campus is residential, which has important advantages. It makes the help lab and working with a partner convenient for all and it provides opportunities for advice when needed from nearby upperclass students. Transporting ZAP! to a different kind of campus may not be easy.

We will be glad to consult with anyone who is interested in adapting the course. We can also supply kits and student books at the cost to the students (around $100, subsidized by the Caltech physics department). Please contact Jerry Pine at jpmail@capsi.caltech.edu

# An Assessment and Evaluation Rubric for a First-Year Engineering Design Optimization Project

**(Article submitted for presentation only)**

**George PLATANITIS and Remon POP-ILIEV**
**Faculty of Engineering and Applied Science, University of Ontario Institute of Technology**
**Oshawa, Ontario L1H 7K4, Canada**

## 1. INTRODUCTION

In the following paper, a rubric for an experimentally supported design project intended to provide first-year engineering students with opportunities to apply robust design and Design of Experiments methodologies is introduced. The project assignment charges students with the task of determining a set of parameter values for an aircraft wing that would optimize its Lift-to-Drag ratio. Students were given background information and necessary equations, leaving them with the task of running the calculations and employing Taguchi methods to select an optimal set of parameter values. General performance on the project has been barely satisfactory, prompting the development of a rubric for students to refer to for project requirements and to guide them in obtaining better grades. The introduced rubric will be implemented in future offerings of this project and its effectiveness evaluated in helping students in achieving better grades.

## 2. BACKGROUND

Design of Experiments has been widely used in industry and in research. For example, DOE was used to improve product and process integrity [1], optimize antibody production [2], and to minimize quality loss when unbalanced tolerances are used in manufacturing settings by identifying a process mean target value [3]. Further, ample scope of DOE applications has been demonstrated in biotechnological applications [4]. As it has become evident that engineers and managers often lack the expertise to use DOE techniques, and are often unaware of their potential benefits, training opportunities have been introduced [5, 6]. At University of Ontario Institute of Technology (UOIT), a decline in performance of students on the optimization project assignment and lowered use of DOE methods in the third-year engineering course from year to year prompted an early introduction of these methods and robust design concepts, in the first-year Engineering Graphics and Design course. As such, a project was introduced in which students were charged with the task of designing an aircraft wing, selecting a parameter combination via DOE, namely via Taguchi methods, that would optimize the Lift-to-Drag ratio (the objective function).

## 3. RUBRIC INTRODUCTION

The development of a rubric to guide students through the requirements of the project and to provide the instructor with an objective method of evaluation was found to be necessary. The proposed two dimensional rubric is based on the ICE (ideas, connections, extensions) methodology [7], rubrics of which are already employed in UOIT's core design courses in engineering design, as well as the final year capstone courses [8, 9]. The proposed requirements for this project to be listed in the rubric include: background search, design considerations, experimental procedure, noise factors, results, and final report. For each requirement, descriptors will be given for each ICE level of application of the requirement corresponding with a respective mark.

## 4. CONCLUSION

The use of rubrics since their introduction in Fall 2007 courses at UOIT has been very successful in providing students guidance as to design assignment/project requirements, as well as providing instructors with a method for fair and objective evaluation of design projects. The rubric proposed herein intends to build on this success by way of providing students with the necessary required components of the wing design optimization project and their levels of application, thereby, giving students the opportunity to maximize their project mark.

## 5. REFERENCES

[1] Antony, J. and Kaye, M., "Design of Experiments for Improving the Product and Process Integrity." *Total Quality Management*, 8(2), 1997, pp. 75-78.

[2] Kallel, H., Zaïri, H., Rourou, S., Essafi, M., Barbouche, R., Dellagi, K., and Fathallah, D. M., "Use of Taguchi's Methods as a Basis to Optimize Hybridoma Cell Line Growth and Antibody Production in a Spinner Flask." *Cytotechnology*, 39(1), 2002, pp. 9-14.

[3] Ming-Hsien, C. L. and Chen, C. J., "Determining Process Mean for Machining while Unbalanced Tolerance Design Occurs," *Journal of Industrial Technology*, 17(1), 2000, 6 pages.

[4] Rao, R.S., Kumar, C.G., Prakasham, R.S., and Hobbs, P.J., "The Taguchi Methodology as a Statistical Tool for Biotechnological Applications: A Critical Appraisal," *Biotechnology Journal,* 3(4), 2008, pp. 510-523.

[5] Antony, J. and Antony, F. J., "Teaching Advanced Statistical Techniques to Industrial Engineers and Business Managers," *Journal of Engineering Design*, 9(1), 1998, pp. 89-100.

[6] Antony, J. and Capon, N., "Teaching Experimental Design Techniques to Industrial Engineers," *International Journal of Engineering Education*, 14(5), 1998, pp. 335-343.

[7] Young, S. F., and Wilson, R. J., *Assessment and Learning: The ICE Approach*, Winnipeg: Portage & Main Press, 2000.

[8] Platanitis, G. and Pop-Iliev, R., "Assessment and Evaluation of Undergraduate Design Engineering Projects," *ASME 2007 International Design Engineering Technical Conference & Computers and Information in Engineering Conference (IDETC/CIE 2007)*, Las Vegas, NV, September 4-7. Paper ID: DETC2007-35383.

[9] Pop-Iliev, R., and Platanitis, G., "Training Well-Equipped Design-Ready Engineering Professionals," *The 9th International Conference on Engineering and Product Design Education (EDPE07)*, Newcastle upon Tyne, United Kingdom, September 13-14, 2007. pp. 129-134.

# A Modern Approach to the Socratic Teaching Method for EEET 2011

**Kelvin WILLIAMS**
**Faculty of Science, Engineering and Computing (SEC), Kingston University, London, KT1 2EE, UK**

## ABSTRACT

An investigation to determine the possibility of using the Socratic Teaching Method (STM) with modern day computer aided learning techniques to deliver lectures in Computer Science to first year undergraduates at university. STM is not defended in any way since its use has long been established. Neither is STM assessed or compared to other teaching methods. The aim of the investigation was simply to find a way of using STM with a modern university Virtual Learning Environment (VLE) – Blackboard V8, in this case. After several attempts over a period of time, a successful way was eventually found and since then, another instructor in the same department has used this method to teach other students on a similar module. This will be of interest to instructors in higher education with an interest in Socratic Teaching.

**Keywords:** Socratic Teaching, Higher Education, Computer Science, Virtual Learning Environment.

## 1. INTRODUCTION

STM is a very old, but tried and tested teaching method. It is a technique of teaching by "asking/discussion" rather than the traditional (Sophic) method of teaching by "telling". It is an interactive process whereby students are stimulated to think about the subject through a series of questions instead of simply being told something in a passive manner. STM is often used on Law courses [1] and postgraduate courses such as MBA [2], but in the author's experience, it is rarely encountered on other courses. In modern times, STM seems to be largely forgotten at undergraduate level and the introduction of new computer aided learning tools have only made it more difficult to use in practice. This is unfortunate, since STM has many advantages over traditional teaching methods.

### 2.1 Disadvantages of the Sophic Teaching Method

Students can sometimes find the traditional Sophic Teaching method uninteresting at times. This is often the case when an instructor talks to the class as a whole, leaving some individuals feeling left-out or ignored. When students lose interest, they often talk amongst themselves and the general background noise level in the class can rise to unacceptable levels. Keeping all the students awake and interested throughout such a lecture can sometimes be difficult using traditional teaching methods.

### 1.2 Advantages of the Socratic Teaching Method

With STM, students are guided towards an outcome by a series of appropriate questions and discussions. The STM process is far more interesting for both student and instructor because students no longer sit in a passive role but they are engaged in an interactive process of questions, answers and discussions. In fact, with STM, the students provide the answers themselves and they find this much more interesting than simply being told something. Student behaviour is better when using STM because the instructor is constantly facing the class and asking questions instead of talking away into space or writing on a board with his or her back to the class. Actually, it is quite common for instructors use STM, but only briefly - to ask a question or two. On such occasions, instructors may observe that students become more attentive and background noise levels drop significantly. What is not so common, however, is for instructors to conduct the entire class session in such a question and answer mode.

## 2. PROBLEMS

Using STM is not a problem in itself. Students are asked a series of questions and they are guided by the instructor to a particular outcome. However, in a modern day university the use of STM is hampered in practice by other factors.

### 2.1 Computer Based Learning Tools

In the past, students attended classes and wrote their own lecture notes but these days, using modern computer based learning tools (Virtual Learning Environments – VLE), lecture notes are provided by instructors and the students can view them using a PC or a laptop. With internet facilities, lecture notes can be accessed remotely and students do not even have to attend classes. Also, with the introduction of high tuition fees, students expect or even demand to have lecture notes provided for them in advance of lectures. As a result of this, many students feel there is no need to take notes in class and they now attend university without pen and paper but armed only with a memory stick to download and store lecture notes that are provided for them. This is acceptable for traditional teaching where students are simply told things but for STM, this poses a problem.

The STM process relies on a system of questions, answers and discussions. The students provide their own answers but the outcomes need to be remembered or written down for future reference. If students are provided with lecture notes in advance, they can see the answers to the questions and the STM experience will be lost on them. Furthermore, if students no longer take notes in class, they will have difficulty remembering the outcomes and main points of lecturers. The problem here is, if students no longer take notes in class then how can STM be used with lecture notes provided in advance without including the answers? One possible solution might be to provide lecture notes with gaps instead of answers so that the student can fill in them in during the class. This would be fine if students had a pen but these days many students do not. Alternatively, students could edit the lecture notes on their laptops but this only works if all students have one.

Although modern computer learning tools have helped to provide students with good quality lecture notes, the classroom experience has become very passive for them – all they have to do is just sit and listen. This is not suitable for STM.

**2.2 Large Classes with Mixed Ability and Mature Students**

In order to meet budgets and boost student numbers, many universities now accept large numbers of unskilled students with little or no previous knowledge in a particular subject. As long as they have sufficient entry level points, students can now apply to study almost anything. Also, with high levels of unemployment, workers from industry are encouraged to retrain for new careers.

On a first year Computer Science course at university, students from a computing background are mixed with students from other backgrounds and mature students from various backgrounds. Some students will have advanced computer skills, some students will have basic computer skills and some students will be much older, have mixed skills and will not have seen a class room for twenty to thirty years.

Teaching large size, first year classes of mixed ability, including mature students is difficult using any teaching methods and the following problems were identified:-

1. Teaching skilled students at a suitable level without making it too difficult for unskilled students.
2. Teaching unskilled students at a suitable level without making it too easy for skilled students.

Lectures must not be too difficult for the unskilled students and yet, at the same time, they must be demanding enough for the skilled students. The problems of teaching such classes affects all teaching methods but the use of STM can be made easier if these problems are resolved beforehand.

## 3. HYPOTHESIS

The problem of mixed ability classes was thought to be solvable by providing lecture notes (aimed at the skilled students) online in advance of the class along with presentation slides (aimed at unskilled students) during the class. In this way, all the students could read the lecture notes in advance and if they could understand them (skilled students), they need not attend the class. However, if some students found the lecture notes too difficult (unskilled students), then they could attend the class where they would receive additional and more detailed tuition using presentations slides. It was hoped that this would reduce the class size to a more manageable number and pitch the work at a level acceptable to both skilled and unskilled students. The problem of teaching the unskilled students with a technical topic without making it too difficult for them could then be solved in class using presentation slides combined with STM - provided the VLE lecture note problem could be solved first.

## 4. METHODOLOGY

STM was introduced gradually over a number of lectures in order to gain confidence with using it fully. To start with, a hybrid version was used and questions were presented on Power Point slides with the answers displayed on the following slide. However, this was not in the true spirit of Socratic Teaching since the answers are supposed to come from the students and not the instructor. Sometimes, there was a need to show a diagram or equation and this was simply inserted into the slide show. Eventually, it was realised that a question could be posted on a slide and following a short discussion, the answer from the students could be typed in beneath the question, during the class. As long as the answers were relatively short there was minimum delay. To avoid complications, a printed set of questions and suitable short answers were taken into the class for reference by the instructor. As far as the students were concerned, it was their answers that were being recorded. At the end of the lecture, all changes to the slide show were saved and the presentation was posted on the VLE for the students to access later. In total, around forty questions were asked and answered in a two hour time slot with a short break midway. The inspiration for using STM was a web site on teaching binary numbers to children in schools [3]. For full details, the Rick Garlikov web site [3] is recommended and instructors can adapt the process to suit their own particular needs.

## 5. RESULTS

STM required some practice to get it right but after a few attempts it worked very well and the lectures ran quite smoothly. Not only did the unskilled students learn something new and technical but they also enjoyed the interactive experience. The skilled students were satisfied with the online notes and they were happy enough to not attend the lectures on topics they were already familiar with. The learning outcomes of the module were exactly the same as those with the traditional teaching method and the only difference was that the material was presented in a slightly different way. With STM the class behaviour was better and the module assessments results were similar to those from previous years (using traditional teaching) - taking into account different students, different questions and different times. As far as student results were concerned, there were no significant differences between the two different teaching methods. From the instructor point of view, STM was certainly a more interesting teaching method but it required more preparation to find a suitable combination of questions and answers for use with the VLE.

## 6. CONCLUSIONS

It was indeed possible to integrate STM with a university VLE to teach first year Computer Science students. STM required more careful planning and preparation than the traditional teaching method but overall, its advantages outweighed its disadvantages – students were taught the same topics with similar results and class behaviour was improved. It is not possible to say if STM could be used on all subject matter, but it was possible to insert diagrams and equations into the Power Point slide show in a non-STM way. Lectures on different topics, could in theory, use a combination of STM and traditional methods where suitable.

Overall, the use of STM in this topic was an enjoyable but demanding experience for the instructor. Something new is always a challenge but it can only serve to make life more interesting. There can be nothing worse than teaching the same thing year after year in the same old way.

## 7. STM POWER POINT SLIDES – BINARY NUMBERS

Available upon request – k.williams@kingston.ac.uk

## 8. ACKNOWLEDGEMENTS

1. Graham Alsop for advice and recommendations
2. Phillip Houghton - for using the STM Power Point slides in another module.

## 9.REFERENCES

1.Kerr, O.S.,(1999) “The Decline of the Socratic Method at Harvard”, 78 Neb L. Rev.,p 113-134
2. Goldratt, E.M., (1997)“Critical Chain, The North River Press, 1997, ISBN 0-88427-153-6
3. Garlikov R.(2006)“The Socratic Method:Teaching by Asking Instead of by Telling”www.garlikov.com/writing.ht


# AUTHORS INDEX
## Volume II
## (Edition Post-Conference)