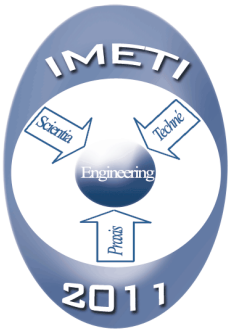


The 4th International Multi-Conference on Engineering and Technological Innovation

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

PROCEEDINGS
Volume I

Edited by:

Nagib Callaos
Hsing-Wei Chu
José Ferrer
Andrés Tremante
C. Dale Zinn

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-978-1-936338-36-8 (Volume I)

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

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	**88**	**100.00**
United States	24	27.27
Japan	10	11.36
France	8	9.09
Italy	7	7.95
Germany	4	4.55
China	3	3.41
Czech Republic	3	3.41
India	3	3.41
Canada	2	2.27
Oman	2	2.27
Poland	2	2.27
Rwanda	2	2.27
Spain	2	2.27
Taiwan	2	2.27
Denmark	1	1.14
Indonesia	1	1.14
Iraq	1	1.14
Ireland	1	1.14
Lithuania	1	1.14
Mexico	1	1.14
Norway	1	1.14
Russian Federation	1	1.14
Saudi Arabia	1	1.14
Slovenia	1	1.14
South Africa	1	1.14
Turkey	1	1.14
Uganda	1	1.14
United Kingdom	1	1.14

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*, 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, EEEP 2011, EEET 2011, and OEPT 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 th 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 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

VOLUME I

CONTENTS

Engineering and Society

Engineering Concepts, Relations and Methodologies

Engineering Education

Materials Sciences and Engineering

Mechanical Engineering including Industrial Engineering, Operations Research, Aerospace, Marine and Agricultural Engineering, Mechatronics, Robotics

Technological Development and Innovation

System Ageing Assessment in Energy Supply Security Model

Juozas AUGUTIS
Department of Mathematics and Statistics, Vytautas Magnus University,
Kaunas, Lithuania

Ričardas KRIKŠTOLAITIS
Department of Mathematics and Statistics, Vytautas Magnus University,
Kaunas, Lithuania

and

Inga ŽUTAUTAITĖ-ŠEPUTIENĖ
Laboratory of Nuclear Installation Safety, Lithuanian Energy Institute
Kaunas, Lithuania

ABSTRACT

This study presents an algorithm developed for the assessment and updating estimates of the parameters in mathematical models of device or system ageing process (that is characterized by an increasing failure rate) with respect to prior information and obtained observations (failure data). The proposed algorithm is based on modified application of Bayesian approach (BA). The paper presents some methods for forecasting the remaining lifetime with respect to the allowed risk margins using the obtained updated mathematical model of increasing failure rate. The developed methodology is used for the modeling of technical disturbances in energy security analysis model [1].

Keywords: ageing, increasing failure rate, Bayesian approach.

1. INTRODUCTION

Security of energy supply has recently gained importance on the policy agenda due to the growing dependence of industrialized economies on energy consumption and the increased frequency of supply disruption. Security of energy supply is a very important field of national security in every country. It includes mining, conversion and transportation of primary energy sources, generation, distribution and supply of energy, functioning of infrastructure, secure life of society from technical, economical, socio-political and environmental point of view. Energy security level [2] is measured by a system of indicators. According to this system, the results of each scenario simulation are transferred into numerical value of security level. Security measure is integral and covers the most important technical, economical, environmental and political criteria of energy supply [2]. This research work is focused on assessing the reliability of technical systems in face of ageing phenomena.

Reliability or efficiency of the system can change (i.e. be decreasing) because of its ageing. The ageing of the system, which could be understood as a general process in which the characteristics of components, systems and structures ("devices") gradually change with time or use, eventually leads to degradation of materials subjected to service conditions and could cause a reduction in component and systems safety margins. Usually some devices or systems can work safely for a longer time than it is denoted in the technical specifications. Moreover, the replacement of devices (or some separate parts) is expensive, costly (in meaning of the time) or could be related to certain risk. In engineering management it is important to determine the maximum remaining lifetime with the allowed risk level.

The process of system ageing can be defined by various characteristics that depend on time (i.e. use, or other factors). The paper analyzes the case when ageing process is characterized by an increasing failure rate that depends on time.

For the solution of the discussed problems the following tasks are going to be analyzed:

- Estimation of the moment at which the ageing process starts (or determining that ageing process has already started).
- Development of an algorithm for obtaining and updating the estimates of parameters in the mathematical models of ageing process with respect to prior information and new observations.
- Forecast of safety duration of the system operation and the remaining lifetime. For instance, determination of point and interval estimates of the critical time moment in respect of the allowed critical value of increasing failure rate.

In the paper, a scheme of modified application of Bayesian approach (BA) that is suitable for updating random parameters estimates in the mathematical model of non-stationary processes as ageing is presented.

2. SYSTEM AGEING PROCESS CHARACTERIZED BY INCREASING FAILURE RATE

Device / system operation time can be divided into three parts (Figure 1): *burn-in* period (when failure rate is decreasing); period of *useful life* (characterized by constant failure rate); and *wear-out* (or so called *ageing*) period (when failure rate is increasing): at the time moment t_0 it starts, the critical value of failure rate marked λ_{cr} is reached at the time moment t_{cr}.

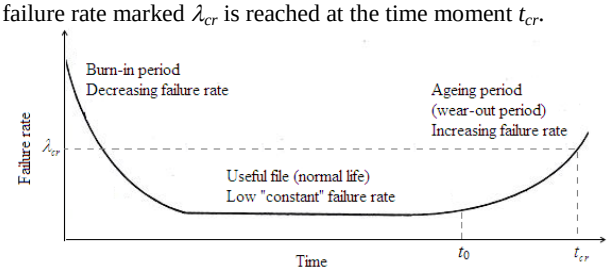


Figure 1. The bathtub curve. Hypothetic failure rate versus time.

It is evident that the dependability of the considered system / device / software is decreasing in the third ageing period because of more and more frequent failures. So it is important to determine whether the ageing period has started for the analyzed system / device; to develop a mathematical model for the system / device ageing period that enables making point and interval forecasts of the remaining lifetime of the device (i.e. failure rate does not exceed the allowed critical value λ_{cr}).

The estimation of the moment at which the ageing process starts

There are some tests for determining whether discrete events in a process have a trend: Laplace test or so called *centroid test* [6], *inversion test* [4, 8], *two-cells test* [3].

Failure rate trend updating

Assume that the system ageing process has already started, and it is characterized by an increasing failure rate. Sometimes the approximate failure rate dependence on time (or other factors) for the particular groups of devices is known in advance. However, prior information can lead to some uncertainties, thus the parameters of this dependence are assumed as random variables.

Lifetime distribution of the considering system is selected according to the form of failure rate trend. Usually the applied lifetime distributions are characterized by increasing failure rate (i.e. "↑"), or bath-tube (i.e. "U") shape failure rate function: Weibull, Birnbaum-Saunders (↑); Generalized Modified Weibull Exponentiated Weibull, Additive Weibull, Modified Weibull, Modified Weibull Extension (↑ and U), etc.

In reliability / dependability analysis of technical device, technical systems, or even software, one of the most popular distribution of failure data is Poisson distribution (with parameter λ). This distribution can be used in case of non-constant failure rate: except that in Poisson model, parameter λ is replaced with $\lambda(t)$, i.e. function of time t, or any other that depends on one (or more) factor(s). In this case it is defined as a mathematically simple model of lifetime. Apart from that, the trend curve of increasing failure rate could be any in respect of the analyzed system or device.

The next step is to evaluate the parameters in lifetime distribution. The maximum likelihood estimation (MLE) is quite popular statistical method used for providing estimates for the model parameters. The limitation of this method is that there is no possibility to use prior information about random parameters; its estimates are obtained using only statistical data. If the distributions of the model parameters are known, they are involved into the model through the law of total probability; in this case the obtained model is complex, and it usage is quite complicated.

The method that allows involving statistical data and prior information about distributions and estimates of random parameters is Bayesian approach; using this method the obtained posterior distribution (parameters estimates as well) could be updated by new available statistical data.

Modified application of Bayesian approach: Commonly Bayesian approach is applied to update the estimated parameters of stationary process when more statistical information becomes available. In case of non-stationary processes analysis, the available statistical data can not be used to update the characteristics of the previous period, because it represents the other state of the system. For analyzing non-stationary process the required information is the following:

- distribution of statistical data;
- form of the trend of system dynamics describing characteristic ξ (as a function of some factors $F_1, \ldots, F_r$ and parameters $\theta_1, \ldots, \theta_s$: $f(\theta_1, \ldots, \theta_s, F_1, \ldots, F_r)$, for example, it is exponential, polynomial, linear, etc. The expected value of random non-stationary characteristic ξ satisfies the requirement

$$\mathrm{E}\xi = f(\theta_1, \ldots, \theta_s, F_1, \ldots, F_r). \quad (1)$$

Note that parameters of distribution of statistical data / observations are expressed in respect of eq. (1) must be satisfied. BA is applied to update random unknown parameters $\theta_1, \ldots, \theta_s$ of the function defined by eq. (1).

In respect of the uncertainty of prior information, the measurement of noise and the parameters $\theta_1, \ldots, \theta_s$ of the considered model could be assumed as random independent variables with priori known distributions (otherwise non-informative, for instance, uniform distribution can be used as prior; note that their prior pdfs are $p_l(x_l)$, $l = 1, \ldots, s$).

Assuming that the distribution of statistical data y_i, $i = 1, \ldots, m$, is also known, i.e. the likelihood function $L(\cdot)$ satisfies eqn (1), the posterior multidimensional density function is obtained by the application of Bayesian formula for this information

$$\varphi(x_1,..., x_s \mid y_1,..., y_i) = \quad (2)$$
$$= \frac{\prod_{l=1}^{s} p_l(x_l) \cdot L(y_1,..., y_i \mid x_1,..., x_s)}{\int_{R_1}\ldots\int_{R_s} \prod_{i=1}^{s} p_l(u_l) \cdot L(y_1,..., y_i \mid u_1,..., u_s)\mathrm{d}u_1 \ldots \mathrm{d}u_s},$$

$i = 2, \ldots, m$, R_l – range (set of all possible values) of parameter θ_l, $l = 1, \ldots, s$.

Some complications arise in practical computation of the integral that appears in the denominator of Bayes formula. Some simplifications for this problem are possible (of instance, using conjugated pairs of prior density function and likelihood function = avoidance of numerical calculation of integral gives convenient application of BA). Depending on the likelihood function, some prior distributions can always lead to the posterior distribution, which has the same functional form as the prior distribution, for instance, Normal likelihood leads to the conjugate posterior Normal distribution. This statistical property is related to the so-called conjugate pair of prior distribution and likelihood [5]. Using the conjugate pairs, the mean and variance as well as other parameters can be easily estimated for the posterior distribution. Using a conjugate pair of likelihood and prior numerical errorless and convenient algorithms for updating, the distributions can be developed [7].

Bayesian point estimate (expected value of posterior distribution) of parameter θ_l is

$$\hat{\theta}_{li} = \int_{R_1}\ldots\int_{R_l}\ldots\int_{R_s} x_l \cdot \varphi(x_1,..., x_l,..., x_s \mid y_1,..., y_i)\mathrm{d}x_1 \ldots \mathrm{d}x_l \ldots \mathrm{d}x_s, \quad (3)$$

$i = 2, \ldots, n$.

The asymptotic of Bayesian point estimate: The main criterion for evaluating the uncertainty of the Bayesian estimate is the analysis of its variance. In general, it is difficult to research it because the variance is expressed as an integral. Only some cases of the convergence of Bayesian estimates are presented (when conjugate pairs [3] are used). For instance, in case of the prior Gamma pdf and Poisson likelihood, the expected value and variance are

$$\mathrm{E}\lambda_n = \frac{b_n}{a_n} = \frac{b + \sum_{i=1}^{n} y_i}{a + n}, \quad \mathrm{Var}\lambda_n = \frac{b_n}{a_n^2} = \frac{b + \sum_{i=1}^{n} y_i}{(a+n)^2} = \frac{\mathrm{E}\lambda_n}{a+n}.$$

If $E\lambda_n \underset{n\to\infty}{\to} const$, $\mathrm{Var}\lambda_n \underset{n\to\infty}{\to} 0$; rate of convergence – $\frac{1}{n}$, here n – amount of statistical data.

Note that in all cases of conjugate pairs, the variance of Bayesian point estimate is decreasing with convergence rate not lower than 1/n. The common convergence of Bayesian point estimate, that is accuracy of updated estimate of random parameter, is still an open problem (if in all cases the variance of updated estimate is decreasing).

Bayesian interval estimate: In systems reliability analysis, the estimates of model parameters are not always sufficient: it is necessary to obtain their confidence intervals (with a given confidence level) as well. It is necessary to construct a modified asymmetric confidence interval, if the beginning (or the ending) of this interval has a higher importance for reliability analysis. For instance, asymmetric interval is a useful interval estimate for λ_{cr}. The definition of this kind of interval is as follows:

Definition 1. Asymmetric confidence interval of the unknown true value of θ for a given significance level α is credible interval $[\hat{x}_1, \hat{x}_2]$ that satisfies these equalities

$$P(x_1 \le \theta \le x_2) = 1-\alpha,$$

$$P(\theta < x_1) = c\alpha,\ P(\theta > x_2) = (1-c)\alpha,\ 0 < c < 1.$$

For instance, the asymmetric confidence interval of Normal random variable (i.e. $N(\mu, \sigma)$) is

$[\mu - u_{c\alpha}\cdot\sigma;\ \mu + u_{(1-c)\alpha}\cdot\sigma]$, here $u_{\cdot}$ – quantile of standard Normal distribution.

Performing the uncertainty analysis it is important to construct the narrowest confidence interval of the considered random parameter. For instance, if the random parameter is the moment at which the system ageing starts, its formal definition may be set as follows:

Definition 2. Credible interval of the unknown true value of θ, for a given significance level α, is the narrowest confidence interval $[\hat{x}_0; \hat{x}_0 + \Delta x]$:

$$\hat{x}_0 = \arg\min_{x_0} \Delta x \text{, when } \Delta x \text{ is defined}$$

$$P(x_0 \le \theta \le x_0 + \Delta x) = 1-\alpha.$$

Note. Credible interval posterior (obtained by BA) pdf is used for the calculation.

Quite often complex calculations of the credible intervals limit their applicability in practice (for instance, in real time decision based algorithms that must be convenient, i.e. not requiring so much time for performing the calculations.

Some aspects of BA application:

- Behavior of the model (for instance, linear or exponential, etc.) does not change when applying BA. BA updates just estimate the parameters in the chosen model.
- BA allows updating the estimates of all parameters in the model with a single new obtained observation.
- This method does not demand to collect observations. Information about the observation is incorporated into the distributions of model random parameters through the likelihood function.

Modified application of BA for NHPP (non-homogenous Poisson process) data: Assume that device failure number per time unit (marked as k) follows Poison distribution with a time dependent parameter $\lambda = \lambda(i, \theta)$, i is interpreted as time, and parameter θ is random variable with known pdf $p(x)$. Then the posterior pdf is obtained using the observations of the system failure data and Bayesian approach

$$p(x \mid k_1, ..., k_n) = \frac{p(x)\cdot\prod_{i=1}^{n}\lambda(i,x)^{k_i}\,\mathrm{e}^{-\lambda(i,x)}}{\int_0^{\infty} p(u)\cdot\prod_{i=1}^{n}\lambda(i,u)^{k_i}\,\mathrm{e}^{-\lambda(i,u)}\mathrm{d}u}. \quad (4)$$

Case 1. Assume that the failure rate trend is linear, i.e. $\lambda(i, \theta) = \theta \cdot i$, the conjugate prior pdf of random parameter θ is Gamma $Ga(a, b)$. Posterior pdf is Gamma $Ga(a_n, b_n)$ as well, with parameters $a_n = a+1+2+...+n$, $b_n = b + \sum_{i=1}^{n} k_i$. Bayesian point estimate of random parameter θ

$$\hat{\theta}_n = \frac{b + \sum_{i=1}^{n} k_i}{a+1+2+...+n}, \quad (5)$$

with variance

$$\mathrm{Var}\hat{\theta}_n = \frac{b + \sum_{i=1}^{n} k_i}{(a+1+2+...+n)^2}. \quad (6)$$

In case of $n+1^{th}$ observation k_{n+1} the parameters of a new posterior pdf could be easily recalculated $a_{n+1} = a_n + (n+1)$ and $b_{n+1} = b_n + k_{n+1}$ estimate of θ as well.

Case 2 (numerical example). Assume that failure rate trend is exponential, i.e. $\lambda(i, \theta_1, \theta_2) = \theta_1 \mathrm{e}^{\theta_2 \cdot i}$, and prior pdf of random parameter θ_2 is exponential with parameter μ. Bayesian point estimate of random parameter θ_2 is (note: parameter θ_1 is assumed to be known)

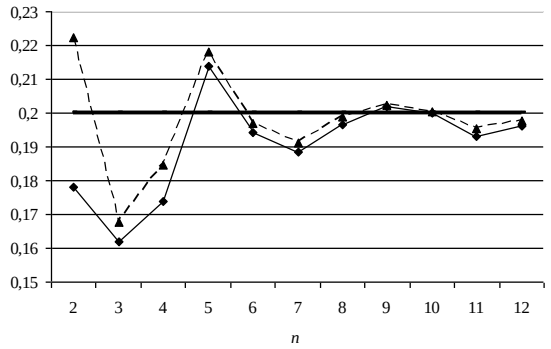
$$\hat{\theta}_{2n} = \frac{\int_0^{\infty} x\cdot \mathrm{e}^{x\left(\mu+\sum_{i=1}^{n} i\cdot k_i\right)}\cdot \mathrm{e}^{-\theta_1\sum_{i=1}^{n}\mathrm{e}^{x\cdot i}}\,dx}{\int_0^{\infty} \mathrm{e}^{x\left(\mu+\sum_{i=1}^{n} i\cdot k_i\right)}\cdot \mathrm{e}^{-\theta_1\sum_{i=1}^{n}\mathrm{e}^{x\cdot i}}\,dx}, \text{ with variance}$$

$$\mathrm{Var}\hat{\theta}_{2n} = \frac{\int_0^{\infty} (x-\hat{\theta}_{2n})^2\cdot \mathrm{e}^{x\left(\mu+\sum_{i=1}^{n} i\cdot k_i\right)}\cdot \mathrm{e}^{-\theta_1\sum_{i=1}^{n}\mathrm{e}^{x\cdot i}}\,dx}{\int_0^{\infty} \mathrm{e}^{x\left(\mu+\sum_{i=1}^{n} i\cdot k_i\right)}\cdot \mathrm{e}^{-\theta_1\sum_{i=1}^{n}\mathrm{e}^{x\cdot i}}\,dx}.$$

A numerical experiment was performed to illustrate the convergence (see Figure 2) of Bayesian point estimate to true value. Failure number k_i, $i = 1, \ldots, 12$, in the i^{th} interval of time is simulated by Poison distribution with parameter $\lambda^*(i) = \theta_1 \mathrm{e}^{\theta_2 \cdot i}$ ($\theta_1 = 5$, $\theta_2 = 0{,}2$). The numerical experiment parameter θ_2 is assumed as random variable (case 1: with a prior non-informative distribution, i.e. its density function is constant; case 2: with a prior informative distribution). For the case 2, an exponential distribution with parameter $\mu = 3$ was chosen as a prior known informative distribution.

An alternative method for the estimation of the parameter in the model with known trend function is least square method. According to the obtained results, the sum of errors squares of BA (with non-informative prior distribution) is approximately 12% bigger than BA (with prior exponential distribution); the sum of errors squares of LSM is approximately twice bigger than BA (with prior exponential distribution). Obviously if a set of observations is quite big, all methods give quite precise estimates of the parameters. BA power, i.e. the combination of

prior information and observations (likelihood as well) is the case of just few observations (i.e. in the beginning).



▲– with exponential distribution as prior pdf (dash line)
♦ – with non-informative prior pdf (solid line)
– – true value of parameter $\theta_2 = 0{,}2$.

Figure 2. Bayesian point estimates of θ_2 ($n = 1, \ldots, 12$).

Random parameter in the trend of failure rate is replaced with the updated Bayesian point estimate (expected value). In the future analysis, it could be used to develop a mathematical method for the assessment of the moment at which the increasing failure rate would exceed the allowed failure rate.

Forecast of the remaining lifetime in respect of increasing failure rate

In the common case, an assumption is made that trend function of increasing failure rate depends on s parameters

$$\lambda = \lambda(t, \theta_1, \ldots, \theta_s), \tag{7}$$

critical value λ_{cr} of increasing failure rate is defined (in technical specifications or determined by experts) for analyzed device or system; its corresponding time moment t_{cr} – remaining lifetime.

This section analyses the problem of evaluating the remaining lifetime of a device and proposes two algorithms.

I. The expected value of the remaining lifetime t_{cr} can be estimated solving the equation

$$\lambda_{cr} = \lambda(t_{kr}, \hat{\theta}_{1n}, \ldots, \hat{\theta}_{sn}), \tag{8}$$

here $\hat{\theta}_{in}$ – Bayesian point estimate of parameter θ_i, $i = 1, \ldots, s$, obtained by formula (3).

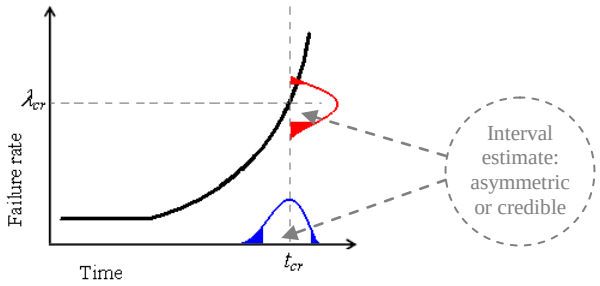


Figure 3. Increasing failure rate in the period of ageing.

II. a) Failure rate $\lambda(t)$ (defined by (7) equation) is a function of random variable(s) $\theta_1, \ldots, \theta_s$, its probability density function (pdf) is obtained using the posterior pdfs of θ_i, $i = 1, \ldots, s$, and transformation formulas (some cases are presented in Table 1).

Table 1. Probability density functions $f(y \mid k_1, \ldots, k_n)$ of failure rate $\lambda(t)$.

Expression of $\lambda(t)$ (t – time)	Random parameter	Pdf of failure rate $\lambda(t)$ (t – time)
$\lambda(t) = \theta \cdot t$	θ	$f(y) = \frac{1}{t} p\left(\frac{y}{t}\right)$
$\lambda(t) = \theta_1 e^{\theta_2 \cdot t}$ (note: θ_1 is known)	θ_2	$f(y) = \frac{1}{\theta_1 t} \frac{1}{y} p\left(\frac{\ln y}{\theta_1 t}\right)$

Note: $p(\cdot) = p(x, k_1, \ldots, k_n)$ – posterior pdf of random parameter defined by formula (4).

For the given fixed time moment t^* the probability that the increasing failure rate would not exceed the critical value λ_{cr} could be easily calculated

$$P(y < \lambda_{cr} \mid t = t^*) = \int_0^{\lambda_{cr}} f(y \mid t = t^*) \mathrm{d}y . \tag{9}$$

II. b) On other hand, it is relevant to evaluate the remaining lifetime of the analyzed device or system with respect to the allowed risk level expressed by the critical value of failure rate. The mean value of the remaining lifetime is not precise enough. In fact, the lifetime t_{cr} is a function of random variable(s) $\theta_1, \ldots, \theta_s$ defined by equation (implicit form)

$$\lambda_{cr} = \lambda(t_{cr}, \theta_1, \ldots, \theta_s) . \tag{10}$$

Pdf of t_{cr} is obtained using the posterior pdfs of θ_i, $i = 1, \ldots, s$, and transformation formulas (some cases are presented in Table 2).

Table 2. Probability density functions $g(z \mid k_1, \ldots, k_n)$ of the remaining lifetime t_{cr}.

Trend function $\lambda(t)$ (t – time)	Lifetime t_{cr} as function of random variable	Pdf of remaining lifetime t_{cr}
$\lambda(t) = \theta \cdot t$ random parameter θ	$t_{cr} = \frac{\lambda_{cr}}{\theta}$	$g(z) = \frac{\lambda_{cr}}{z^2} p\left(\frac{\lambda_{cr}}{z}\right)$
$\lambda(t) = \theta_1 e^{\theta_2 \cdot t}$ random parameter θ_2 (note: θ_1 is known)	$t_{cr} = \ln\left(\frac{\lambda_{cr}}{\theta_1}\right) \frac{1}{\theta_2}$	$g(z) = \ln\left(\frac{\lambda_{cr}}{\theta_1}\right) \frac{1}{z^2} p\left(\ln\left(\frac{\lambda_{cr}}{\theta_1}\right) \frac{1}{z}\right)$

Note: $p(\cdot) = p(x, k_1, \ldots, k_n)$ – posterior pdf of random parameter defined by formula (4).

Pdf $g(z)$ that contains information about the observations and prior distributions of random parameters gives point and interval estimates of lifetime t_{cr}. In the point of interval estimate, the construction of asymmetric (with more attention given to the beginning, see Figure 3) or credible interval has more advantages than classical confidence interval with equal-tails. For instance, credible interval is the shortest of all confidence intervals and contains more probable value of the remaining lifetime.

3. RESULTS AND CONCLUSIONS

The algorithm for the estimation of the time moment, at which the ageing process of device or system starts, was developed (it was based on testing of parametric hypothesis with respect to failure data).

The paper presented the algorithm for obtaining (and updating) the estimates of random parameters in the mathematical models of characteristics that describe system, device or substance ageing: the behavior (trend function dependent on some factors and parameter(s)) of the analyzed characteristics is priory prior known. The proposed algorithm is based on the modified application of Bayesian approach. In the paper, the illustration of the developed algorithm applicability was presented by a numerical experiment: in case when ageing process is characterized by an increasing failure rate with prior known trend; the convergence of Bayesian point estimate of the parameter of failure rate trend function was demonstrated; the obtained results were compared with the ones calculated using least squares method (LSM): the sum of errors squares of LSM was approximately twice bigger than the sum of errors squares of BA.

The algorithms for the assessment point and interval estimates of the device or system remaining lifetime (i.e. forecast of safety duration of the device or system operation) were presented in case when the allowed risk level was determined using critical value of increasing failure rate.

The developed methodology is used for technical disturbances modeling in energy security analysis model.

ACKNOWLEDGMENTS

This research was funded by a grant (No. ATE-08/2010 and ATE-10/2010) from the Research Council of Lithuania.

4. REFERENCES

[1] Augutis J., Matuzienė V., Krikštolaitis R., Pečiulytė S., Norvaiša E. 2008. Analysis of security of energy supply assessment methods. Energetika., Vol. 54(4), 1-9.

[2] Augutis J., Krikštolaitis R., Pečiulytė S., Konstantinavičiūtė I. 2011. Sustainable Development and Energy Security Level After Ignalina NPP Shutdown. Technological and Economic Development of Economy, Vol. 17(1), 5-21.

[3] Atwood C., Cronval O., Patrik M., Rodionov A. 2007. Models and data used for assessing the ageing of systems, structures and components (European network on use of probabilistic safety assessment (PSA) for evaluation of ageing effects to the safety of energy facilities). EUR 22483 EN. Petten: EC DG JRC Institute for Energy.

[4] Bendat J. S., Piersol A. G. 1986. Random data: analysis and measurement procedures. New York: Wiley.

[5] Bernardo J. M., Smith A. F. M. 2003. Bayesian theory. John Wiley & Sons.

[6] Birolini A. 2007. Reliability Engineering. Theory and Practice. ISBN 978-3-540-49388-4 5th.ed.Springer Berlin Heidelberg New York, 323-326.

[7] Littlewood B., Popov P., Strigini L. 2000. Assessment of the Reliability of Fault-Tolerant Software: a Bayesian Approach. Proc. 19th International Conference on Computer Safety, Reliability and Security, SAFECOMP'2000, Rotterdam, the Netherlands, Springer.

[8] Rodionova A, Atwood C L, Kirchsteiger C, Patrik M 2008. Demonstration of statistical approaches to identify component's ageing by operational data analysis—A case study for the ageing PSA network. Reliability Engineering and System Safety, vol. 93 1534-1542.

DNA Separation by Surface Electrophoresis

Avital Braiman
Division of Engineering, Brown University,
Providence, RI 02912, USA

Center for Engineering Science Advanced Research,
Computer Science and Mathematics Division,
Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

Thomas Thundat
Department of Chemical and Materials Engineering,
University of Alberta, Edmonton, AB T6G 2V4, Canada

Fedor Rudakov
Center for Engineering Science Advanced Research,
Computer Science and Mathematics Division,
Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

DNA separation by size is required for a wide variety of applications including DNA sequencing, genotyping, mutation analysis, STR forensic analysis, and diagnosis of disease. Separation is usually achieved by electrophoretic transport of charged biomolecules through a sieving media such as gel[1], polymer solution[2,3], or microfabricated sieves[4]. Since separation effectiveness of sieving media based techniques is determined by the size of the sieve relative to the size of the DNA fragments to be sorted, sieving media based techniques are generally effective in a limited range of DNA sizes. Therefore, it is necessary to prepare separate sieving matrices for each DNA size range of interest.

Recently, it was experimentally demonstrated that DNA molecules can be separated by size by electrophoretic transport across a planar or a nanopatterned substrate[5,6,7]. This approach is different from other commonly employed approaches in that it does not utilize a sieving medium but rather relies on frictional interactions of the DNA molecules with the surface. As a consequence, it does not have the constraints of sieving media based techniques and can be applied for simultaneous separation of DNA fragments ranging in size from hundreds of basepairs to several megabase pairs.

DNA separation mechanisms on surfaces are not fully understood. While the power law behavior of the DNA mobility with the DNA size has been experimentally demonstrated, understanding the cause of this behavior remained an intriguing and unsolved problem. We present a bead-spring model utilizing a minimal set of parameters that provides insight into the mechanisms of DNA separation on surfaces[8]. Our results elucidate the experimentally observed power law trend of the DNA mobility with the DNA size N ($\mu \approx N^{-\alpha}$)[8]. Previous attempts to explain DNA dynamics on the surface assumed that separation is caused by different degrees of adsorption of short and long DNA fragments. In our model we assumed that DNA fragments have the same degrees of adsorption regardless of their sizes and separation results due to excitation of propagating wave modes[8].

Our simulations show that for electric fields just slightly exceeding the minimal field required to start motion, the DNA mobility is approximately inversely proportional to its size (the scaling exponent α is very close to 1). For larger values of the electric field, the DNA mobility scales as $N^{-\alpha}$ with $0 < \alpha < 1$. In addition, the mobility trend shows an oscillatory behavior. At large values of the electric field, the DNA mobility becomes size independent (α=0).

[1] J. L. Viovy, "Electrophoresis of DNA and other polyelectrolytes: physical mechanisms" **Reviews of Modern Physics,** Vol. 72, No. 3, 2000, pp. 813-872.
[2] G. W. Slater, M. Kenward, L. C. McCormick, and M. G. Gauthier, "The theory of DNA separation by capillary electrophoresis" **Current Opinion in Biotechnology,** Vol. 14, issue 1, 2003, pp. 58-64.
[3] K. D. Altria, "Overview of capillary electrophoresis and capillary electrochromatography" **Journal of Chromatography A,** Vol. 856, Issus 1-2, 1999, pp. 443-463.
[4] W. D. Volkmuth and R. H. Austin, "DNA electrophoresis in microlithographic arrays" **Nature**, Vol. 358, 1992, pp. 600-602.
[5] N. Pernodet, V. Samuilov, K. Shin, J. Sokolov, M. H. Rafailovich, D. Gersappe, and B. Chu, "DNA electrophoresis on a flat surface" **Physical Review Letters,** Vol. 85, Issue 26, 2000, pp. 5651-5654.

[6] Y.-S. Seo, V. A. Samuilov, J. Sokolov, M. Rafailovich, B. Tinland, J. Kim, and B. Chu, "DNA separation at a liquid-solid interface" **Electrophoresis,** Vol. 23, Issue 16, 2002, pp. 2618-2625.
[7] Y.-S. Seo, H. Luo, V. A. Samuilov, M. H. Rafailovich, J. Sokolov, D. Gersappe, and B. Chu, "DNA electrophoresis on nanopatterned surfaces" **Nano Letters**, Vol. 4, No. 4, 2004, pp. 659-664.
[8] A. Braiman, T. Thundat, F. Rudakov, "DNA separation on surfaces" **Applied Physics Letters,** Vol. 97, Issue 3, 2010, pp. 033703/1-3.

Study of an RFId smart label for food quality and safety

Giuseppe PERRETTI, Gian Franco REGNICOLI
Department of Economic and Food Science, Section of Food Technology and Biotechnology
University of Perugia
Via S. Costanzo n.c.n., 06126 - Perugia, Italy
E-mail: perretti@unipg.it

Marco GRASSI, Piero MALCOVATI
Department of Electrical Engineering, University of Pavia
Via Ferrata 1, 27100 - Pavia, Italy

ABSTRACT

Nowadays, the quality and food safety assurance is a prerequisite from which we cannot escape. Many documents in matters of food law and recent facts, like several scandals, confirm this.
For this purposes, the implementation of evolved and low-cost tracing and tracking systems based on micro-electronic and miniaturized devices, wich are able to identify the food products, to track the travelled routes and to oversee their preservation conditions, could be proposed.
In this context, the aim of this paper is to investigate the possibility to improve food quality and safety, focusing on the study and the application of a smart label equipped with innovative sensors, hardware and software, useful for the measurement and the control of environmental parameters for the shelf-life of foodstuffs and for their identification by the radio frequency technology (active RFId).
The laboratory experiments, with test of the proposed microsystem on measuring temperature and light intensity with the accuracy of 0.5 °C and 6 W/m^2 respectively, confirm how this integrated multi-sensor technology can be considered for the improvement of food quality and safety. The experimental models suggest the storage and transmission possibility of such data in case of need, by the use of the proposed smart RFId label. Actually, tests on ham and bread products are in progress, and preliminary results gave the positive opinion that there are possibilities of improvement and the study of the productive chains.

Keywords: RFID, smart label, food quality, food safety, traceability.

1. INTRODUCTION

Today, as underlined by the last scandals and facts, the food safety assurance, considered as "the prevention of illnesses resulting from the consumption of contaminated food" [1] is a prerequisite from which we cannot escape. In confirmation of this, there are several documents in matters of food law, since the European legislation: first of all the "Food Safety White Book" [2], then the European Commission Regulation 178/2002 [3] about food safety, which instituted the European Food Safety Authority (EFSA), then the European set of laws about hygiene of 2004 (Regulation CE n. 852, 853, 854, 882/2004 and Directive CE 2002/99), and finally the voluntary ISO 22000:2005 concerning food safety management systems [4, 5-8].
The importance of this topic is widespread in all the world. In the USA the Food and Drug Administration (FDA) has recently developed a comprehensive Food Protection Plan and a set of food safety laws to address the changes in food sources, production, and consumption [9]. Many are the studies of food safety assurance systems in regions like China [10], Australia and New Zealand [11]. These kinds of studies are extended also on many developing countries [12-13].
Nevertheless, despite the several mandatory and voluntary measures for food safety, some difficulties still exist in realizing such efforts, as, for example, the lack of a common language able to allow the diffusion and sharing of information at all the levels of the food chain.
This situation leads up the operators to the choice from several useful systems to track the product in the internal process. Sometimes these systems are not each-other fully compatible.
A smart identification system, should also be a helpful instrument for prevention toward the food hazards, because it allows to make focused recalls,

with the goal to fix the issue just on the suspected items [14].

Compared with a massive recall, the possibility to have a right definition of the batches brings some advantages, like a save of time, energy and resources that mean a lower economic impact and a higher probability to realize an effective recall.

In recent years, consumers have been frequently exposed to food alerts (from adulterations and sophistications to contamination of pesticides and dioxin, till the recent danger of the *Escherichia Coli* case in Germany) that could have had a more reduced impact, if Tracking and Tracing (T&T) and safety tools had worked more effectively [9].

During the development of the investigation, concerning quality and safety control, hardware instruments and innovative applications were evaluated (e.g. RFID – Radio Frequency Identification – tags over integrated multi-sensory data). These will be able to get and record information related to environmental parameters, such as for example time, temperature and humidity of the packaged atmosphere, useful to control the chemical composition changes that influence the hygienic, nutritional and sensory characteristics of foodstuffs.

Literature shows the importance of Innovative hardware tools for the improving of food quality and safety [15-18].

Among this tools, RFID (Radio Frequency Identification) are transponders able to recognize in the distance objects, animals and people by the use of radio waves. In recent years, this mature technology has received significant coverage, mainly because of the considerable use of this technology made by private companies such as Wal-Mart (USA), Tesco (UK), by the Department of Defence (USA) and proposed by the Food and Drug Administration (FDA) [19].

In the literature, many potential benefits from the utilization of RFID tags are described [20].

Among the tools that promise to help food safety and quality, we can find the "active" and "intelligent" packaging. The active packaging is defined by The European Parliament and The Council of the European Union, [21-22] as packaging that "deliberately incorporate 'active' components intended to be released into the food or to absorb substances from the food" and changes the condition of the packed food to extend shelf-life or to improve safety or sensory properties, while maintaining the quality of packaged food. Examples of active packaging are the O_2 scavengers, the humidity or SO_2 emitters, the ethylene or CO_2 scavengers and emitters and the films with antimicrobial activity [23]. Instead, intelligent packaging has been defined as "articles are designed to monitor the condition of the food", as stated by Kerry et al. [23], "packaging systems which monitor the condition of packaged foods to give information about their quality during transport and storage". In this category are included the Time Temperature Indicators (TTI), the indicators of oxygen, the CO_2 indicators and the microbial growth indicators [23].

TTIs are presented as labels on the food packaging. They are easy to understand by the consumer and have a large use for packaging of fresh milk, frozen fish, meat and seafood [24].

Nevertheless, nanotechnology is the science dealing with particles of size from 1 to 100 nm. It is an interesting approach that in recent times is having a strong development and attention. These containers are more resistant to tears and shocks, or more impermeable against gases, such as water vapour, or oxygen and UV radiation [25].

Literature shows that, in the area of food safety, with particular need for T&T, the interviewed actors are current and prove susceptible to RFID technology. In some cases the actors are already studying applications of this technology on food production [26].

Another significant aspect is the ability to transfer to the consumer the benefits above mentioned, because customers are able to associate with relative ease using sensors / tags to a guarantee of safety and better quality of the marketed product [18,28].

2. OBJECTIVES

In the actual context, characterized by a high attention on food quality and safety [29], it is interesting to explore the applicability of the above presented innovative technologies for the food safety and quality. There is a greater sensitivity toward the innovative sensors able to track quality indices of the shelf life such as temperature, pH, humidity inside the packaging, and, for particular preparations such as intermediate humidity bakery packed products (snacks), even the ethanol concentration in the atmosphere inside the package.

In fact, the implementation of microelectronic systems for tracing food products through low-cost miniaturized devices is quite important. These devices shall be able to identify the product, track its path along the complete food distribution chain, and monitor the environmental conditions to which the product is subject over time, in order to verify that the preservation prescriptions are respected, for example, to define a dynamic expiration date and avoid the distribution of degraded products [29-30].

In this context, the first step toward the implementation of the smart label is an integrated

micro-system which includes sensors and very low power circuits (sensor interface circuits and A/D converter). Humidity, temperature and light intensity represent the most interesting quantities to qualify the preservation environment of food products, as well as those that realistically can be measured with a low-cost and low power integrated circuit [31].

The goal of this study was to investigate the possibility to improve food quality and safety, focusing on the study and the application of a smart label equipped with innovative sensors, hardware and software, useful for the measurement and the control of environmental parameters for the shelf-life of foodstuffs and for their identification by the radio frequency technology (active RFId).

With this aim, laboratory tests were carried out, in order to analyze the hardware and software technology useful for this project. The results of these activities are here reported.

3. RESULTS

The fabricated chip, whose microphotograph is shown in **Fig. 1**, occupies an area equal to 4 mm^2 including I/O pads.

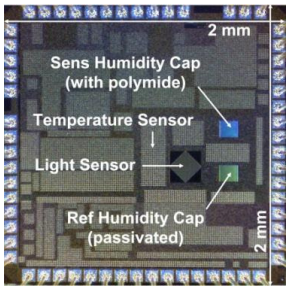


Fig.1: Microphotograph of the developed integrated device.

The humidity sensor is realized using an interdigitated capacitor coated with polyimide, whose dielectric permittivity changes with the relative humidity (RH) [32-33]. The light intensity sensor consists of an integrated reverse-biased photodiode, which delivers a current proportional to the incident light intensity.

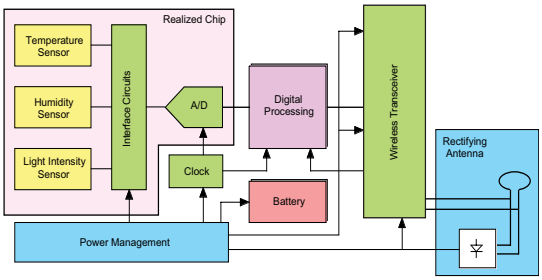


Fig.2: Block diagram of the presented smart label system

Humidity sensor	
Range, resolution	10% ÷ 90% RH, 1%
Maximum current consumption	20 µA
Temperature sensor	
Range, resolution	-1 °C ÷ 75 °C, 0.2 °C
Maximum current consumption	30 µA
Light intensity sensor	
Range, resolution	1000 W÷m^2, 2 W÷m^2
Maximum current consumption	30 µA
A/D converter	
Input range peak-peak differential	1 V
Resolution	10 bits
Maximum current consumption	20 µA
RFID transceiver	
Operating frequency	13.56 MHz
Standard	ISO 15693
Maximum current consumption	150 µA
Rectifying antenna	
Output voltage, current @ recharging	3.3 V, 100 µA
Output voltage, current @ transmission	1.8 V, 150 µA
Power management	
Regulated output voltage	1.8 V
Output current	200 µA

Tab. 1: Specifications of the proposed smart label

In **tab. 1** the specifications of the smart label are reported. The large temperature range resolution (-1 ÷ 75 °C) allows the device to be suitable for measurement of different food productions. Other specifications are in previous work of the same authors [29].

The light intensity is a critical parameter for the preservation of foods sensible to oxidation, like fats or oil containing foods, or for foods that after exposure of light can rise to off-flavours, for example beers or wines.

Each of the three considered sensors (humidity, temperature, and light intensity) is connected to its specific interface circuit, which delivers a voltage output suitable for A/D conversion. One single A/D converter is used for all the sensors, which are therefore read-out sequentially, through a multiplexer. When a specific sensor is not used, it is switched-off in order to reduce the power consumption. The A/D converter is also powered down when idle. The frequency of the measurements and, hence, the average power consumption can be customized at system level, depending on the specific use of the device.

Finally, **Fig. 3** and **Fig. 4** show the measurement results of the response of the complete acquisition chain in controlled environment to temperature and light intensity respectively. We obtained better

than 0.5 °C accuracy in the [-1 °C, +75 °C] temperature range and better than 6 W/m^2 in the [5 W/m^2, 320 W/m^2] light intensity interval.

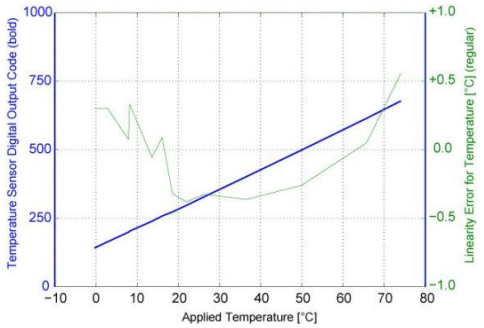


Fig. 3: Temperature sensor read-out response (bold) and linearity error curve (regular) in controlled environment.

In **table 3** the Temperature sensor read-out response (bold) and linearity error curve (regular) in controlled environment are reported. The temperature sensor, exploits the difference between the base emitter voltages of two bipolar transistors with different emitter areas, to produce a voltage proportional to the absolute temperature [34].

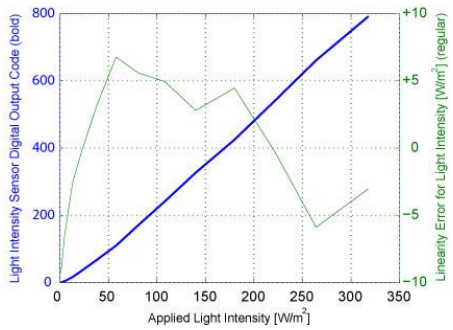


Fig. 4: Light intensity sensor read-out response (bold) and linearity error curve (regular) in controlled environment.

The light intensity sensor consists of an integrated reverse-biased photodiode, which delivers a current proportional to the incident light intensity.

4. CONCLUSIONS

In this study the possibility to improve food quality and safety, focusing on the study of a smart label equipped with innovative sensors, hardware and software has been investigated. Such smart label is, useful for the measurement and the control of environmental parameters for the shelf-life of foodstuffs and for their identification by the radio frequency technology (active RFId).
At the actual state of development of the device, temperature, and light intensity responses were tested, with results that confirm the aptitude of the sensors to the realization of a smart label. The results of the preliminary tests on food (data not reported) and the results above described, are encouraging for the next activities, that will regard the test of the humidity sensor and relative application of the sensors on food products.

5. AKNOWLEDGEMENTS

This study was realized with the found of the **Italian Ministry of University and Research,** project n° RBOP06AMPP.

6. REFERENCES

1) R. Akkerman, P. Farahani, M. Grunow, "Quality, safety and sustainability in food distribution: a review of quantitative operations management approaches and challenges", **OR Spectrum**, 32, 2010, pp. 863–904.
2) Commission of The European Communities: "**White Paper On Food Safety**", Brussels, 12 January 2000.
3) The European Parliament and the Council of the European Union, "Regulation (EC) No. 178/2002", **Official Journal of the European Communities**, 1.2.2002, L31/1-L31/24. /2002.
4) Council Directive 2002/99/EC of 16 December 2002, **Official Journal of the European Union**, 23.1.2003, L 18/11.
5) The European Parliament and the Council of the European Union, "Regulation (EC) No 852/2004", **Official Journal of the European Uni**on, 30.4.2004, L 139/1.
6) The European Parliament and the Council of the European Union, "Regulation (EC) No 853/2004", **Official Journal of the European Union**, 30.4.2004, L 139/55.
7) The European Parliament and the Council of the European Union, "Regulation (EC) No 854/2004", **Official Journal of the European Union**, 25.6.2004, L 226/83.
8) The European Parliament and the Council of the European Union, "Regulation (EC) No 882/2004", **Official Journal of the European Union**, 30.4.2004, L 165/1.
9) Nachay K., "New Food Safety Law Addresses Product Traceability" **Food Technology**, vol. 65, No.2, 2011, pp. 44-45.
10) L. Bai, C. Ma, S. Gong, Y. Yang, "Food safety assurance systems in China", **Food Control**, No. 18, 2007, pp. 480-484.
11) M. Healy, S. Brooke-Taylor, P. Liehne "Reform of food regulation in Australia and New Zealand", **Food Control**, No. 14, 2003, pp. 357-365.
12) Kabwit Nguz "Assessing food safety system in sub-Saharan countries: An overview of key issues", **Food Control**, No.18, 2007, pp. 131-134.

13) Tjaart W. Schillhorn van Veen "International trade and food safety in developing countries", **Food Cont**rol, No. 16, 2005, pp. 491-496.
14) S. Kumar, E. M. Budin, "Prevention and management of product recalls in the processed food industry: a case study based on an exporter's perspective". **Technoviation**, No. 26, 007, pp. 739-750.
15) G.F. Regnicoli, P. Malcovati, G. Perretti, "Innovative systems for the improvement of food quality and safety", Proc. of "The 2nd **International Multi-conference on Engineering and Technological Innovation: IMETI 2009**" July 10th -13th, 2009 – Orlando, Florida, Usa.
16) A. Ciani, M.L. Voros, "Agro-Food Traceability Organization (Territorial) Marketing And Movie Oriented A.T.O.M.O", **2nd International Conference on Engineering and Meta-Engineering: ICEME 2011**, March 27th-30th 2011, Orlando, Florida, USA.
17) D. Barling, R. Sharpe, T. Lang, "Traceability and ethical concerns in the UK wheat-bread chain: from food safety to provenance to transparency", **International Journal of Agricultural Sustainibility**, Vol. 7, No. 4, 2009 , pp. 261-278.
18) A. Dalziel Traceability & food: go for the record. **Asia Pacific Industry**, Nov./Dec. 2010, pp. 60-63.
19) T.A. McMeekin, J. Baranyu, J. Bowman, P. Dalgaard, M. Kirk, T. Ross, S. Scmid, M.H. Zwietering, "Information systems in food safety management", **International Journal of Food Microbiology**, No. 112, 2006, pp. 181-194.
20) M. Tajima, "Strategic value for RFID in supply chain management", Jour**nal of Purchasing & Supply Management**, No. 13, 2007, pp. 261-273.
21) The European Parliament and the Council of the European Union, "Regulation (EC) No 1935/2004" **Official Journal of the European Communities**, 13.11.2004, L 338/4.
22) The European Parliament and the Council of the European Union, "Regulation (EC) No 450/2009", 13.11.2004, L 338/4.
23) J.P. Kerry, M.N. O'Grady, S.A. Hogan, "Past, current and potential utilization of active and intelligent packaging systems for meat and muscle-based products: A review", **Meat science**, No. 74, 2006, pp. 13-130.
24) S. Yan, C. Huawei, Z. Limin, R. Fazheng, Z. Luda, Z. Hengtao, "Development and characterization of a new amylase type time-temperature indicator", **Food Control**, No. 19, 2008, pp. 315-319.
25) A. Sorrentino, G. Gorrasi, V. Vittoria, "Potential perspectives of bionanocomposites for food packaging applications", **Trends in Food Science & Technology** , No.18, 2007, pp.84-85.
26) A. Regattieri, M. Gamberi, R. Mancini, "Traceability of food products: General framework and experimental evidence", **Journal of Food Engineering**, No. 81, 2007, pp. 347-356.
27) W. Van Rijswijk, L.J. Frewer, D. Menozzi, G. Faioli, "Consumer perceptions of traceability: A crossnational comparison of the associated benefits", **Food Quality and Preference**, No. 19, 2008, pp. 452-464.
28) Chassy B. M., "Food safety risks and consumer health" **New Biotechnology** Vol. 27, No. 5, 2010, pp. 534-544.
29) D. Cartasegna, A. Cito, F. Conso, A. Donida, M. Grassi, L. Malvasi, G. Rescio, P. Malcovati, "Smart RFID label for monitoring the preservation conditions of Food", **2009 IEEE International Symposium on Circuits and Systems: ISCAS 2009**, May 24th-27th, 2009 Taipei International Convention Center, Taipei, Taiwan.
30) Kelepouris T., Pramatari K., Doukidis G., "RFID-enabled traceability in the food supply chain" **Industrial Management and Data Systems**, 107 (2), 2007, pp. 183-200.
31) Jian Zhang, Lu Liu, Weisong Mu, Liliana M Moga, Xiaoshuan Zhang, "Development of temperature-managed traceability system for frozen and chilled food during storage and transportation" **Journal of Food, Agriculture & Environment**, Vol. 7, 2009, pp. 28-31.
32) T. Boltshauser, L. Chandran, H. Balks, F. Bose, D. Steiner, "Humidity sensing properties and electrical permittivity of new photosensitive polyimides", **Sensors and Actuators B**, No. 5 , 1991, pp. 161-164.
33) T. Boltshaljser, H. Baltes, "Capacitive Humidity Sensors in SACMOS Technology with Moisture Absorbing Photosensitive Polyimide", **Sensors and Actuators** A, 25-27, 1991, pp. 509-512.
34) P. Malcovati, F. Maloberti, C. Fiocchi, M. Pruzzi, "Curvature-compensated BiCMOS bandgap with 1-V supply voltage," **IEEE Journal Of Solid-State Circuits**, vol. 36, No. 7, 2001, pp. 1076–1081.

Low-frequency magnetostrictive inertial actuator

F. Braghin, S. Cinquemani, F. Resta

Mechanical Engineering Department, Politecnico di Milano,
Campus Bovisa Sud, via La Masa 34, 20156, Milano, Italy

ABSTRACT

The present paper presents the design of an innovative low-frequency magnetostriuctive inertial actuator that is able to multiply by almost a factor of 10 the amplitude of striction-elongation amplitude of the magnetostrictive bar, thus leading to an increase in the generated force amplitude, and to obtain a working frequency from 30Hz up, i.e. well below the working frequencies of traditional magnetostrictive inertial actuators. Moreover, the design has been optimized through an analitical model and a finite element model taking into account all design parameters. The optimised low-frequency magnetostriuctive inertial actuator has then be produced and its frequency response compared to that of a traditional magnetostrictive actuator made up of the same components (except for the supporting structure).

1. INTRODUCTION

Magnetostrictive inertial actuators are able to effectively generate an actuation force only above their first eigenfrequency that is determined by the inertial mass and the stiffness along the actuation direction ([1]). In traditional magnetostrictive actuators the inertial mass is directly connected to the actuator foot through the bar of magnetostrictive material (plus a rigid spring that is required for applying the necessary mechanical pre-load to the magnetostrictive bar, [2]). Thus, the stiffness along the actuation direction is strictly related to Young's modulus of the magnetostrictive material. To obtain low-frequency actuators one could therefore either increase the inertial mass (e.g. by adding more weight on top of the actuator) or decrease the cross-section of the magnetostrictive bar (thus reducing its axial stiffness). There is however a limit to the increase in inertial mass and/or decrease in the bar's cross-section imposed by Euler's critical load (buckling of the magnetostrictive bar) and by the frailty of the actuator (magnetostrictive materials usually have a fragile behaviour, [3]). This leads to traditional magnetostrictive actuators that have their first eigenfrequency well above 100Hz ([4]).

To obtain a low-frequency magnetostrictive inertial actuator (with the first eigenfrequency below 100Hz) it is necessary to re-design the actuator in order to decouple the stiffness along the actuation direction from Young's modulus of the magnetostrictive material. In the present paper an innovative design of the magnetostrictive actuator is proposed that uses the same components of traditional magnetostrictive actuators but with a different layout: the magnetostrictive bar is mounted orthogonally with respect to the direction of actuation and is connected to the inertial mass through a special deformable structure having low stiffness along the direction of actuation. This innovative configuration allows to improve two performances:

- the up and down motion of the inertial mass is not equal to the amplitude of striction-elongation (about 1% of the length of the magnetostrictive bar) but can be multiplied by almost a factor of 10 thus leading to an increase in the force amplitude that the inertial actuator is able to generate (at equal inertial mass);

- the eigenfrequency of the actuator is determined by the deformability of the supporting structure and not to the axial stiffness of the magnetostrictive bar thus allowing to reduce the working frequency from about 30Hz on (i.e. significantly extending the bandwidth of the inertial actuator towards low frequencies).

Moreover, the proposed configuration allows to greatly reduce the overall height of the actuator thus allowing to mount it even in narrow cavities.

The design of the low-frequency magnetostrictive inertial actuator has been optimized both through analytical models and through finite element models in order to achieve the best compromise between force amplitude, bandwidth, stresses in the supporting structure, height and complexity of the assembly/final cost of the device. The comparison between the developed models (simple analytical and finite element ones) shows that the analytical model is able to correctly estimate the motion amplification of the innovative device while the finite element model is required for correctly assessing to increase in bandwidth and for determining the stresses the supporting structure will be subjected to. Optimization parameters are all the design parameters of the supporting structure, i.e. its lengths, widths, radii, points of connection to the inertial mass and to the magnetostrictive bar, etc. Also the material of the supporting structure has been considered in this optimization phase although, at the end, the most bounding requirement was that of the cost of the material of the supporting structure.

Once optimized, the low-frequency magnetostrictive inertial actuator has been produced (the supporting structure has been obtained through water-jet technology, and its performances compared to those of a traditional magnetostrictive inertial actuator made up of the same components (except for the supporting structure) and having the same inertial mass ([5]). This comparison has shown to correctness of the design approach and that, through the addition of just the deformable structure, performance indexes are greatly improved.

2. DESIGN

As already observed, magnetostrictive inertial actuators are not able to generate a significant force below their first (axial) eigenfrequency ([1]) and this eigenfrequency is well abobe 100Hz due to the fact that their stiffness is associated to the magnetostrictive bar:

$$\omega_0 = \sqrt{\frac{k_s}{m}} = \sqrt{\frac{A}{s^H L m}} \tag{1}$$

m being the inertial mass fixed on top of the magnetostrictive bar and k_s being its axial stiffness equal to $A/s^H L$ (where $A = 9mm^2$ is the cross-section area of the magnetostrictive bar, $s^H = 3.3 * 10^{-11}\ m^2/N$ its mechanical compliance when a constant magnetic field is applied and L its length). Figure1 shows the eigenfrequency of the magnetostrictive inertial actuator evaluated through equation 1 as a function of the suspended mass and the length of the magnetostrictive bar. It can be seen that, to obtain a magnetostrictive inertial actuator working at low frequencies, a very slender bar with high suspended mass is necessary. However, such configuration is impossible to obtain practically due to the reaching of Euler's critical load (buckling of the magnetostrictive bar).

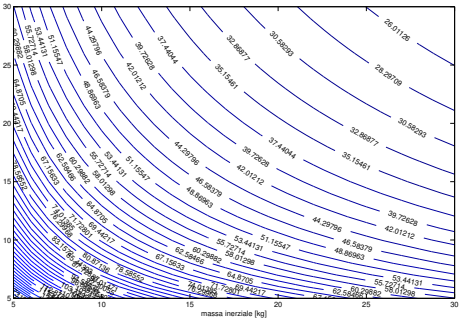


Figure 1: Eigenfrequency of a magnetostrictive inertial actuator as a function of the suspended mass m and bar length L (A=$9mm^2$).

In order to develop a low-frequency magnetostrictive inertial actuator, it is therefore necessary to design a mechanism that is able to decouple the stiffness of the actuator from the stiffness of the axial magnetostrictive bar and, if possible, to amplify the deformations of the bar that are usually about 1% of the bar length itself. The power balance of suche decoupled magnetostrictive inertial actuator sounds:

$$m\ddot{y} \cdot \dot{y} + k_s x \cdot \dot{x} = 0 \tag{2}$$

where x is the elongation of the terfenol-D bar and y is the displacemnt of the inertial mass. The transmission ratio is equal to the ration between the output speed, i.e. $\dot{y}$, and the the input speed, i.e. $\dot{x}$:

$$\tau = \frac{\dot{y}}{\dot{x}} \tag{3}$$

Substituting the definition of τ in equation 2 we get:

$$m\ddot{x} \cdot \dot{x}\tau^2 + kx \cdot \dot{x} = 0 \tag{4}$$

and thus the eigenfrequency of the system becomes:

$$\omega = \sqrt{\frac{k}{m}\tau^{-2}} = \frac{\omega_0}{\tau} \tag{5}$$

It is therefore clear that, by conveniently designing the support structure of the magnetostrictive bar, the eigenfrequency of the actuator can be reduced by a factor equal to the transmission ratio τ.

The proposed mechanism exploits a double motion amplification as schematically shown in figure 2: both the lever mechanism and elastic deformation of the arc/ crossbow cooperate to amplify the deformation of the magnetostrictive bar.

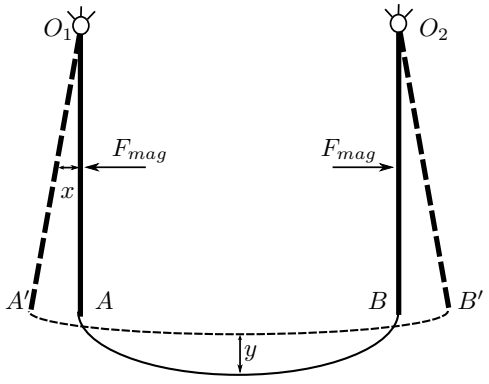


Figure 2: Kinematic scheme of the motion amplification mechanism

3. OPTIMIZATION

In order to design a low-frequency magnetostrictive inertial actuator that is able to maximise the transmission ratio τ as well as the force transmitted tio the inertial mass keeping the stresses of the supporting structure as low as possible, thus guaranteeing an infinite fatigue life, several different configurations were taken into account. Figure 3 shows the main configurations considered.

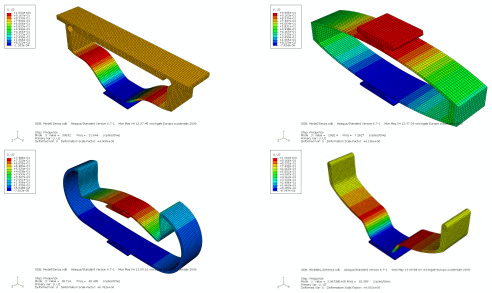


Figure 3: Main configurations considered for the low-frequency magnetostrictive inertial actuator

For each considered configuration, the lever effect, the generated force, the eigenfrequency and the maximum stresses in the supporting structure have been deteremined. From a preliminary analysis, it has been found that the most promising, as well as the easiest one to produce, configuration is the last one shown in figure 3. The parameter optimization procedure has therefore been applied to this last configuration.

At first, a simple kinematic model has been adopted for the optimization procedure to limit the parameter space taking into account just the lever effect. Then, a detailed finite element model has been adopted in order to maximise not only the lever effect but also the generated force as well as to minimize the eigenfrequency and the maximum stresses in the supporting structure.

Analytical Model

In figure 4 the undeformed kinematic model adopted is shown while in figure **??** the deformed kinematic model adopted is siplayed.

The boundary conditions of the analytical model are the following:

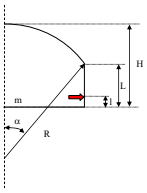


Figure 4: Undeformed configuration of the analytical kinematic model adopted

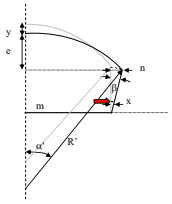


Figure 5: Deformed configuration of the analytical kinematic model adopted

- the maximum dimensions of the support structure (H e m) are fixed,
- the point of application of the displacemnet imposed by the magnetostrictive bar (l) is fixed,
- the amplitude of half the elonfation of the magnetostrictive bar (x) is imposed,
- the length of the arc of the crossbow remains constant ($R\alpha = R'\alpha'$)

The equations that govern the kinematics of the analytical model are the following:

$$\begin{aligned} \alpha &= a\sin\left(\frac{m}{R}\right) \\ L &= H - R\left(1-\cos\alpha\right) \\ \beta &= a\sin\left(\frac{x}{l}\right) \\ n &= L\sin\left(\beta\right) \\ R\alpha &= R'\alpha' \\ m+n &= R'\sin\alpha' \\ e &= R'\left(1-\cos\alpha'\right) \\ y &= H - L\cos\left(\beta\right) - e \end{aligned} \tag{6}$$

The reference configuration of the supporting structure is characterized by a height $H = 15mm$, by a width $2m = 30mm$ and by a point of application of the displacemnet imposed by the magnetostrictive bar $l = 3mm$ from the basis.

The model results are shown in figure 7. As can be clearly seen, the crossbow lever effect (equal to the transmission ratio τ) significantly varies as a function of the crossbow curvature radius and reaches a maximum value of 7.4 for a curvature radius equal to 39.2mm.

Also the effect of other parameters has on the lever effect has been investigated. Figure **??** shows the influence of the width m of the crossbow: at increasing width, higher lever effects are reached but for greater crossbow curvature radii (with $2m = 40mm$ the lever effect becomes equal to 8.3 for a curvature radius equal to 60.9mm).

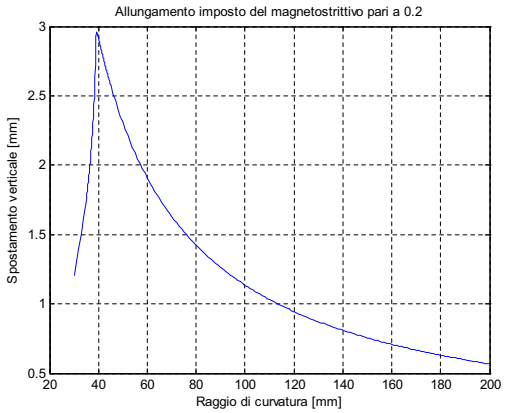


Figure 6: Vertical displacement y as a function of the crossbow curvature radius

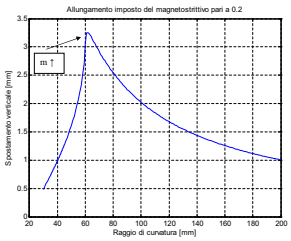


Figure 7: Vertical displacement y as a function of the crossbow curvature radius. Effect of the crossbow width m ($H = 15mm$, $m = 20mm$, $l = 3mm$)

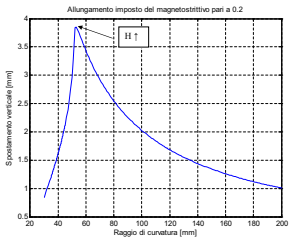


Figure 8: Vertical displacement y as a function of the crossbow curvature radius. Effect of the crossbow height H ($H = 20mm$, $m = 20mm$, $l = 4mm$)

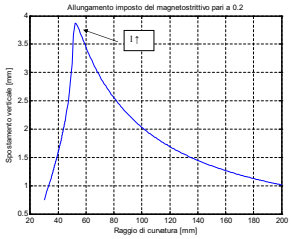


Figure 9: Vertical displacement y as a function of the crossbow curvature radius. Effect of the distance between the hinge and the point of application of the imposed displacement l ($H = 15mm$, $m = 20mm$, $l = 2mm$)

Reducing the distance between the hinge and the point of application of the imposed displacement l (figure 9) a significant increase in the lever effect for slightly greater curvature radii is obtained (with $l = 2mm$ the lever effect reaches a value of 9.5 for a curvature radius equal to 52.2mm). Figure 8, instead, shows how the increase in the height H of the crossbow determines a significant increase in the lever effect (with only a slight increase in the crossbow curvature radii): for $H = 20mm$ the lever ratio becomes 9.2 while the curvature radius becomes equal to 53.0mm (for a distance between the hinge and the point of application of the imposed displacement $l = 4mm$).

Finite Element Model

The optimization of the geometry of the spproting structure as thus been carried out using a detailed finite element model that allows to predict not only the lever effect but also the reaction forces, the eignefrequency of the structure as well as the stresses inside the structure. The developed finite element model of the structure has the following characteristics:

- approx. 7000 linear 8-node brick elements (characteristcis dimensions of the mesh equal to 0.5mm),
- approx. 25000 degrees of freedom,
- lower surface bounded to the ground,
- inertial mass equal to $0.3kg$,
- imposed deformation fo the magnetostrictive bar and
- linear elastic material for the supproting structure ($\rho = 7800kg/m^3$, $E = 210000MPa$, $\nu = 0.3$).

Figures 10, 11 and 12 show, respectively, the vertical displacement y of the inertial mass (in mm) for an elengation of the magnetostrictive bar x equal to 0.4mm, the inertia force generated by the magnetostrictive actuator (in N) for the same elongation of the magnestostrictive bar considered before and the first (vertical) eigenfrequency of the actuator (in Hz) as a function of the crossbow curvature radius and width. Comparing the vertical displacement y obtained from the analytical model and the one determined through the finite element model, a good agreement can be found. Moreover, for a curvature radius of 50mm and a crossbow width of 0.3mm the maximum lever effect is reached. The force generated by the magnetostrictive actuator, instead, does not significantly vary in the parameter range considered (it varies from 370N a 420N) and increases at increasing crossbow width and curvature radius. Finally, the eigenfrequency decreases with the crossbow width and increases with the curvature radius. For the configuration that maximizes the lever ratio,the eigenfrequency is lower than 60Hz. It has therefore been decided to adopt the configuration that maximizes the lever ratio.

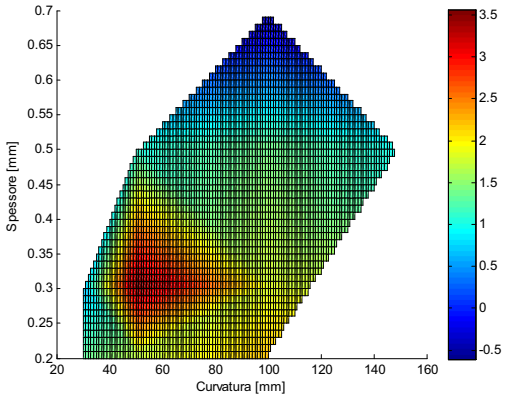


Figure 10: Vertical displacement y of the inertial mass (in mm) for an elengation of the magnetostrictive bar x equal to 0.4mm as a function of the curvature radius and corssbow width

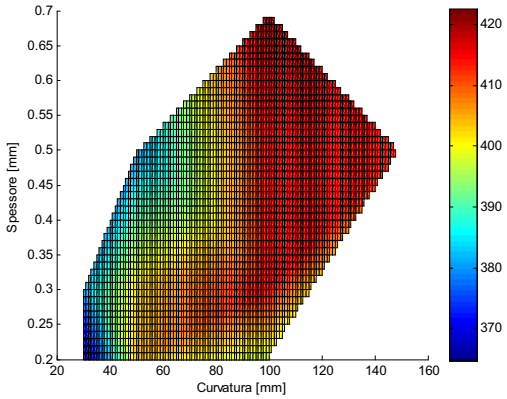


Figure 11: Inertia force generated by the magnetostrictive actuator (in N) for an elengation of the magnetostrictive bar x equal to 0.4mm as a function of the curvature radius and corssbow width

4. PRODUCTION

Figure 13 shows the section of the 3D model of the low-frequency magnetostrictive inertial actuator prototype. The central part is the magnetostrictive bar and at its esxtremities two Neodymius-Iron-Boron magnet discs are places. The supporting structure also supports the windings that generate the magnetic field inside the magnetostrictive material. Table 1 reports the most significant quantities of the designed low-frequency magnetostrictive inertial actuator.

Figure 14 shows a photograph of the produced prototype.

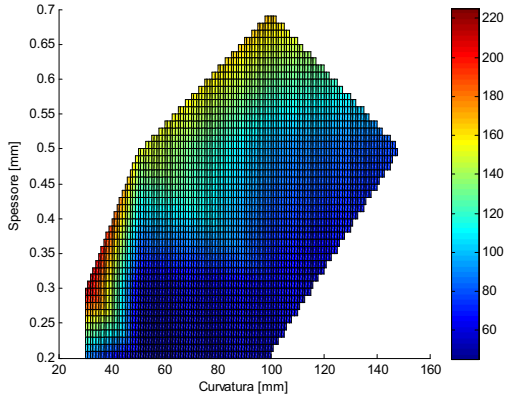


Figure 12: First (vertical) eigenfrequency of the actuator (in Hz) as a function of the curvature radius and corssbow width

Table 1: Most significant quantities of the designed low-frequency magnetostrictive inertial actuator

Quantity	Value
Inertial mass	220 g
Material of supporting structure & mass	steel
Material of magnetostrictive bar	Terfenol-D
Length of magnetostrictive bar	20 mm
Stiffness of magnetostrictive bar	$1.4{\cdot}10^7$ N/m
Number of windings	400
Total height	25 mm
External diameter	54 mm

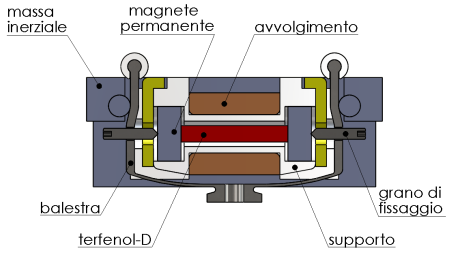


Figure 13: Section of the 3D model of the low-frequency magnetostrictive inertial actuator

5. EXPERIMENTAL RESULTS

The test bench adopted for assessing the low-frequency magnetostrictive inertial actuator prototype characteristics

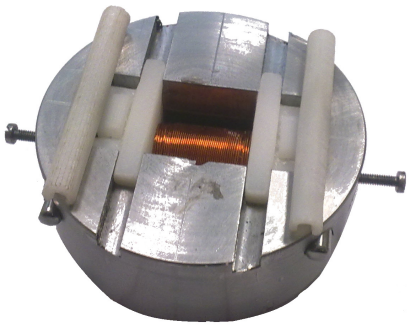


Figure 14: Low-frequency magnetostrictive inertial actuator prototype

is shown in figure 15.

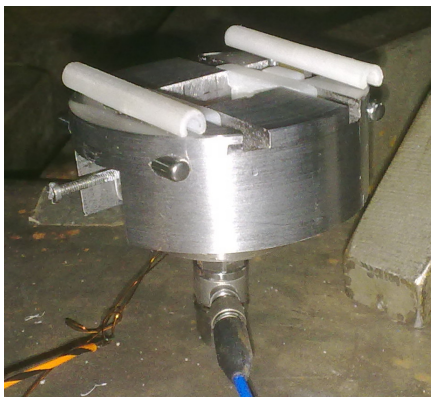


Figure 15: Set up sperimentale

The inertial mass is free to vibrate and the actuator is driven through a time varying supply current I. Both random excitation and sweep sine excitation in the frequency range 1-2000 Hz were adopted. During the test both supply current and voltage are acquired as well as the base and the inertial mass accelerations (through piezolectric accelerometers) and the force transmitted to the ground (through a piezolectric load cell). Figure 16 shows the measured transfer function between supply current I and force transmitted to the ground F_T. It can be seen that the first (vertical) eignefrequency is equal to $\omega = 167Hz$ well below the eigenfrequency of the magnetostrictive inertial actuator without the designed supporting structure ($\omega_0 = 1260Hz$). Moreover, the measured transmission ratio is equal to 7.5 quite close to the expected one.

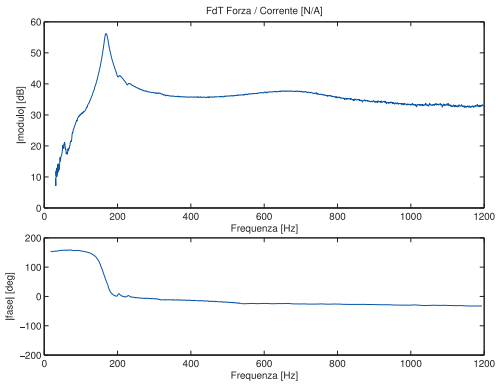


Figure 16: Experimental transfer function $G = \dfrac{F_T}{I}$

6. CONCLUSIONS

In the present paper an innovative supporting structure for obtaining a low-frequency magnetostrictive inertial actuator has been designed and optimised trhough a simple kinematic analytical model and through a complex finite element model. Thus, a prototype has been produced and tested showing a good agreement with the simulated results.

REFERENCES

[1] Braghin F., Cinquemani S., Resta F., "A model of magnetostrictive actuators for active vibration control " *Sensors and Actuators A: Physical* **165(2)**, 342-350, (2011)

[2] Engdahl G., "Handbook of Giant Magnetostrictive Materials", Academic press, (2000).

[3] Clark A.E., Savage H.T., Spano M.L., "Effect of Stress on Magnetostriction and Magnetization of Single Crystal of Tb27Dy73Fe2", *IEEE Transactions on magnetics*, **20(5)**, 1443-1445, (1984).

[4] Dhilsha K.R., Markandeyulu G., Subrahmanyeswara Rao B.V.P., Rama Rao K.V.S., "Design and fabrication of a low frequency giant magnetostrictive transducer" *Journal of Alloys and Compounds*, **258**, 53-55, (1997).

[5] Pons J.L., "Emerging Actuator Technologies: A Micromechatronic Approach", John Wiley and Sons Ltd, (2005).

Design of an actuation system for a variable pitch axial fan

Hermes GIBERTI, Luca PRATO, Ferruccio RESTA

Mechanical Engineering Department, Politecnico di Milano,
Campus Bovisa Sud, via La Masa 1, 20156, Milano, Italy

ABSTRACT

This paper investigates the mechanical design of an axial fan with automatically controlled variable pitch blades, focusing on kinematic and dynamic synthesis of the fan. In addiction it is presented a case of an axial-flow fan with 16 adjustable blades able to settle the air flow rate in a power plant. The dimension of the external hub of the fan is 3.2 [m] and its maximum work velocity up to 1000 rmp.

keywords: axial fun, variable pitch blade.

1. INTRODUCTION

In the last few years the energy problem has become more and more relevant. Besides to the research and the development of new energy sources, the optimization of the existing tools has became a strategic element, especially the optimization of performance of fluid machine. Many examples of axial-flow fans (single or multi stage design) with variable blade pitch angle can be found in energy production field such as in thermal power stations: the two induced-draft fans at the Weiher and Bexbach power stations are among the largest in the world [1] [2].

Fluid-Dynamic analysis of axial-flow fans has been studied for about 30 years [3] [4] [6]. The decision to develop this type of fan was prompted not only by its easily integrable design into overall plant configurations but also and primarily by the operating cost benefits that it offers, specifically when compared with centrifugal fans with variable inlet vanes [5].

This paper investigates the mechanical design of this kind of fans focusing with the kinematic and dynamic synthesis of variable pitch axial flow fans. Different mechanisms to achieve blades regulation are compared and the effects on the overall design of the fan are evaluated. The mechanical design of the fan and the actuation system is considered, neglecting the aerodynamic aspects which are considered only to define the forces applied to each blade. We present a case of an axial-flow fan with 16 adjustable blades to settle the rate of flow. The fan operates as an air supplier system in combustion chamber for energy production. The most relevant operating conditions regard the angular speed 1000 rmp, the starting time of the fan (less than 10 s) and the blade pitch regulation ($\pm 30°$).

Great attention has been paid to the design of the actuation system. It is able to move the blade quickly and accurately, therefore the position of the actuator and the type of supply are evaluated. Moreover, the mechanisms which transform the motion of the actuator in the blade rotation are investigated. An important issue is represented by the choice of blade-shaft bearings and their position, since the centrifugal force can reach more than 50 tonnes for blade. Once the best solution to assembly each blade-shaft on the impeller and to allow an easy maintenance to be made periodically is defined, it is studied the balancing of the blade-shafts by means of appropriate counterweights disposition which is aimed to reduce inertial effects on the system.

Finally in order to compute the transmitted forces and then to size the components, a multi-body model of the fan has been created. This model allows to investigate the problem of balancing blade-shaft and the influence of friction on bearings and aerodynamic forces in the determination of loads on the system. Numerical simulations have been carried out to support the design and highlight project criticalities.

2. ACTUATION SYSTEM: KINEMATIC SYNTHESIS

The pitch blade regulation of an axial-flow fan can be realized with different solution, which can be classified according to the way the motion of the actuator is transformed into rotation of the blades. Each blade is connected to the impeller, so they rotate with it; the blade actuation system can be joined to the impeller or to the ground. Depending on this choice, we can describe these situations:

- in the case the actuator is connected to the impeller, the problem of suppling the actuation system must be solved;
- in the case the actuator is connected to the ground, the relative velocity between the actuator system and the impeller must be added up to zero.

If we choose to have the actuator rigidly linked with the impeller, the force between the impeller and the mechanism to rotate the blades is internal; in this case, it is easier to design the structure and balance the fan, since the forces of the mechanism are not transmitted to the basement. The motion of the actuator can be linear or rotative; in the first case it is necessary to create a mechanism to

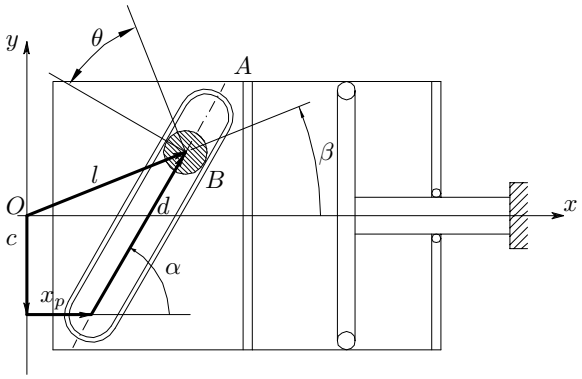


Figure 1: Cinematic draft of *plate and cam configuration*.

convert the linear motion of the actuator in the rotation of the blades. From this analysis, it is clear that two possible kinematic solutions can be used: a rotative actuator linked to the impeller (in this case a motor for each blade or a system of gears is needed) or a linear actuator rigidly constrained to the impeller and a mechanism to convert the motion. Using a motor for each blade, the system can be subject to failure; while, using a system of gears to move all blade with only one motor, it is necessary remove the backslash and plenty lubricate.

The simplest and most suitable choice for this application is a linear actuator with a system to convert the motion. The available technologies for the actuation are electrical, pneumatic or oil-hydraulic. Pneumatic and electromechanical systems play virtually no role in power plant fan engineering, while mechanical control systems of blade pitch used to be employed specifically on smaller units; oil-hydraulic control systems have emerged as the most suitable solution for this purpose. They operate with less hysteresis since they use fewer mechanical power transmission elements; in addition, they are capable of transmitting higher actuating forces. The oil-hydraulic solution grants greater dynamic performances and a better rate weight-power than the other two solutions. The chosen solution for the design of the fan is then a linear oil-hydraulic actuator rigidly linked to the impeller with a mechanism to convert the motion.

Mechanism of motion conversion

The choice of the mechanism of motion conversion is important to determinate the amplitude of the forces between the blade-shaft and the actuator; moreover it defines the relationship between the regulation parameter of the actuator and the angular position of the blade. For this purpose, a mechanism which behaves as a speed reducer (and therefore of position) could minimize the load on the actuator and at the same time it can modulate the action of the actuator itself.

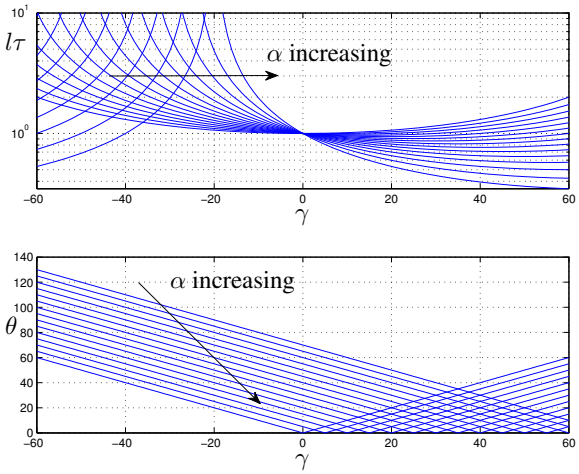


Figure 2: Absolute value of the generalized transmission rate and pressure angle.

The simplest system to convert a linear motion (actuator) to a rotative motion (blade) is obtained by a Scotch yoke mechanism. The piston is directly coupled to a sliding yoke with a slot that engages a pin on the rotating part. In figure 1 it is shown a generic kinematic draft. The cursor A which is rigidly linked to the oil-hydraulic cylinder imposes a rotation to the crank which is connected to the blade-shaft through a prismatic link B. The direction which the point B can move determines if the mechanism is centered ($\alpha = 90°$) or deflected one ($\alpha \neq 90°$). If the Scotch yoke mechanism is centered, the shape of the motion of the link l is constant over the time given a pure sine wave speed of the piston. With reference to figure 1, we can describe the following kinematic relations:

$$x_p = le^{i\beta} - ci - de^{i\alpha} \tag{1}$$

$$\dot{x}_p = i\dot{\beta}le^{i\beta} - \dot{d}e^{i\alpha} \tag{2}$$

$$\ddot{x_p} = il\ddot{\beta}e^{i\beta} - l\dot{\beta}^2e^{i\beta} - \ddot{d}e^{i\alpha} \tag{3}$$

where O is the rotational axis of the blade, x_p is the advance direction of the plate, l is the length of the crank and β is the rotation angle of the blade.

From the relation (2) it is possible to derive the generalized transmission rate τ, that is the rate between the angular velocity of the blade and the translational piston speed:

$$\dot{x_p}\left[\frac{-\sin\alpha}{l\cos(\alpha-\beta)}\right] = \dot{\beta} \quad \rightarrow \quad \tau = \frac{-\sin\alpha}{l\cos(\alpha-\beta)} \tag{4}$$

In the upper graphic of the figure 2, the Absolute value of the generalized transmission rate τ adimensionalized with length l is plotted versus $\gamma = (90° - \beta)$ for $20° < \alpha < 90°$; this graph using a base 10 logarithmic scale

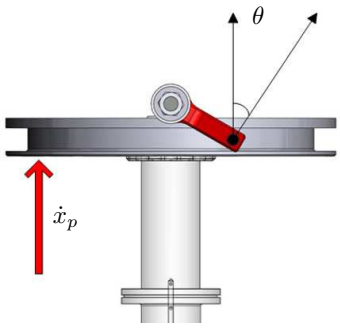


Figure 3: *Plate mechanism.*

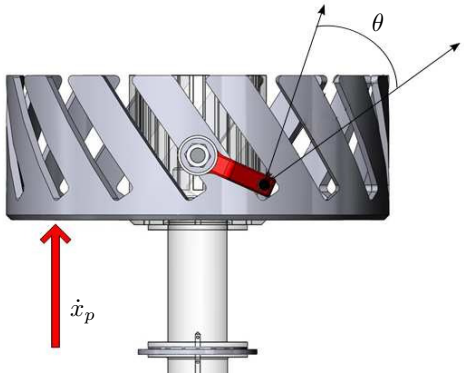


Figure 4: *Cam mechanism.*

for the y-axis. In the lower graph the absolute value of the pressure angle is plotted versus γ for the same previous configurations. The pressure angle is given simply by: $\theta = \alpha - \beta$. In figure 2 the influence of the angle α on the transmission rate and therefore on the pressure angle can be observed.

Figure 3 shows the kinematic solution with the centered Scotch yoke mechanism which has been designed for this fan. The blade-axis is designed with a crank at the end of which a small wheel is mounted. This wheel is constrained between two plates which can translates along the rotational axis of the impeller; this solution is called *plate mechanism* because two plates, by a crank, guides the blade shaft position. The angle θ in figure 3 is the pressure angle.

In order to realize the kinematic solution with a uncentered Scotch yoke mechanism, the crank linked to the blade-shaft is moved through the contact with a conjugate profile, as shown in figure 4. In the proposed solution, the conjugate profile is simply a straight line with a slope of $150°$ from the rotational axis of the impeller. It can be observed, during the design phase, the cam profile can be modified to obtain a variable transmission rate as in the case of spatial cams. For this reason this solution is called *cam mechanism*.

We can immediately observe that with the plate solution the blade rotation has to be symmetrical with respect to the position $\beta = 90°$ and it is impossible to have a rotation greater than $60°$ without worsening the pressure angle. The cam solution turns out to be more flexible and it is possible to set the global rotation $\Delta\beta$ without limitations for symmetry. Moreover the maximum achievable rotation can be increased defining the profile shape. On the other side, the cam solution is more complex to be realized, as shown in figure 4, and the complete mechanism is more cumbersome.

Because of the dimension of the fan and the number of blades, the *plate solution* has been preferred. It can be observed that the *cam mechanism* is more suitable for fans with high number of blades, since the cranks would be aligned with the rotational axis of the impeller, reducing the lateral dimensions that they would have during the motion.

3. BLADE DYNAMIC BALANCING

The chosen solution leads each blade-axis to be unbalanced, with regards to its axis of rotation, due to the presence of the crank: it is impossible to create a crack which is opposite to the command one, since it would interfere with the command crank of the adjacent blade axis. As consequence the bearings are subjected to loads and torques which tend to rotate the blade. To reduce this two effects, first of all it is necessary to move the center of mass on the rotational axes of the blade; in this way it is possible to align a principal axis of inertia with the rotational axes of the blade. The second topic we can focus on regards differences between the other two inertial axis which are responsible for a torque around the rotational axes. To compute this torque it is possible to integrate the effect of each point P of the blade, which is characterized by a mass of ρdV, where ρ is the density of the blade and dV is the infinitesimal volume. The velocity of the generic point P is:

$$\underline{v_p} = \underline{\dot{r}_0} - [\hat{r}]\,\underline{\omega} \tag{5}$$

where $\underline{\dot{r}_0}$ is the velocity of the center of coordinate system "0", against which we compute the moments of inertia, $\underline{\omega}$ is the angular velocity of the point P and $[\hat{r}]$ is the matrix which contains the position of the point P in the reference system "0", matrix $[\hat{r}]$ is defined as:

$$[\hat{r}] = \begin{bmatrix} 0 & -r_z & r_y \\ r_z & -r_y & -r_x \\ -r_y & -r_x & 0 \end{bmatrix} \tag{6}$$

Since the body is rigid, the terms of the matrix$[\hat{r}]$ are constant. Forces and torques are calculated through the inte-

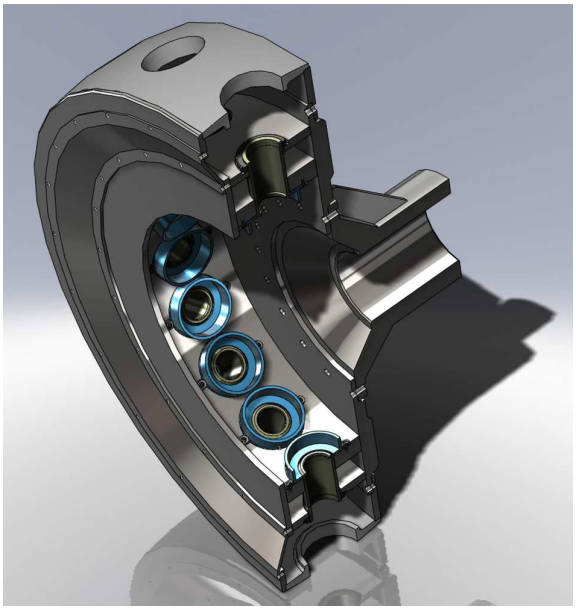


Figure 5: The impeller.

gration of infinitesimal element of mass ρdV, multiplied by the acceleration a_P of each point P; thus it results:

$$\left\{ \begin{array}{c} \underline{F} \\ \underline{C} \end{array} \right\} = \left[\begin{array}{c} \int_V \rho \underline{a}_p \\ \int_V \rho \left[\hat{r}\right] \underline{a}_p \end{array} \right] \tag{7}$$

These can also be written as:

$$\left\{ \begin{array}{c} \underline{F} \\ \underline{C} \end{array} \right\} = \left[\begin{array}{cc} [M] & [S] \\ [S]^T & [J] \end{array} \right] \left\{ \begin{array}{c} \underline{\ddot{r}}_0 \\ \underline{\dot{\omega}} \end{array} \right\} + \left\{ \begin{array}{c} [\hat{\omega}]\,[S]\,\underline{\omega} \\ [\hat{\omega}]\,[J]\,\underline{\omega} \end{array} \right\} \tag{8}$$

Referring forces and torques to the center of mass, one gets:

$$\left\{ \begin{array}{c} \underline{F} = [M]\,\{\underline{\ddot{r}}_{CG}\} \\ \underline{C} = [J]\,\{\underline{\dot{\omega}}\} - [\hat{\underline{\omega}}]\,[J]\,\underline{\omega} \end{array} \right\} \tag{9}$$

$$= \left[\begin{array}{cc} [M] & [S] \\ [S]^T & [J] \end{array} \right] \left\{ \begin{array}{c} \underline{\ddot{r}}_0 \\ \underline{\dot{\omega}} \end{array} \right\} + \left\{ \begin{array}{c} [\hat{\underline{\omega}}]\,[S]\,\underline{\omega} \\ [\hat{\underline{\omega}}]\,[J]\,\underline{\omega} \end{array} \right\} \tag{10}$$

Considering the matrix $[J]$ diagonal and one of the principal axises of inertia (z^*) aligned with the rotational axis of the blade (z), can be defined the angle α as the angle between the principal axis (x^*) and the rotational axis of the impeller (x). Calling $\dot{\Theta}$ the rotational velocity of the impeller around axis x and $\dot{\Psi}$ the rotational velocity of the blade around axis z.

The angular velocity vector of the blade-shaft body is:

$$\underline{\omega} = \left[\dot{\Theta} cos\alpha, \quad -\dot{\Theta} sin\alpha, \quad 0\right] \tag{11}$$

The vector of torques applied in the center of mass is:

$$\underline{C} = -\,[J]\,\{\underline{\dot{\omega}}\} - [\hat{\omega}]\,[J]\,\{\underline{\omega}\{ \tag{12}$$

where:

$$[\hat{\omega}] = \left[\begin{array}{ccc} 0 & -\omega_z & \omega_y \\ \omega_z & -\omega_y & -\omega_x \\ -\omega_y & -\omega_x & 0 \end{array} \right]$$

Deriving $\underline{\omega}$ and replacing $[J]$ with the diagonal matrix of inertia:

$$\left[\begin{array}{ccc} I_{x^*} & 0 & 0 \\ 0 & I_{y^*} & 0 \\ 0 & 0 & I_{z^*} \end{array} \right] \tag{13}$$

one gets: $C_z = \dot{\Theta}^2 \sin\alpha \cos\alpha (I_{x^*} - I_{y^*})$ and $C_x = Cy = 0$. The difference between the two moments of inertia of the blade in the rotational plane of the blade itself generates a torque that makes the blade rotate. This torque tries to align the greater moment of inertia between I_{x^*} and I_{y^*} along the rotation axis of the impeller. Because of the complex geometry of the blade-shaft, the definition of the balancing mass is not easy; it is necessary to identify suitable areas to place the counterweights and to define their mass with numerical optimization algorithms.

4. MECHANICAL SYSTEM DESIGN

Once the kinematic solution and the actuation system have been chosen, the mechanical design of the axial fan can be discussed. The designed solution includes an impeller made of three concentric rings. Each of them is provided with holes, in which the blade-shaft are inserted, as shown in figure 5.

The inner ring is the most critical component, because it has to bear all the centrifugal load of the blades. Moreover the inner ring together with the intermediate one has to bear the radial loads which are responsible of keeping the blade shaft in its bearing housing. The external surface of the third ring has a spherical shape with a diameter of 1920 [mm] letting the blade to rotate without interfering with the impeller and keeping the distance from the base of the blade constant.

Figure 6 shows the designed blade-axis. The blade is constrained to the shaft with six screws M20. On the same

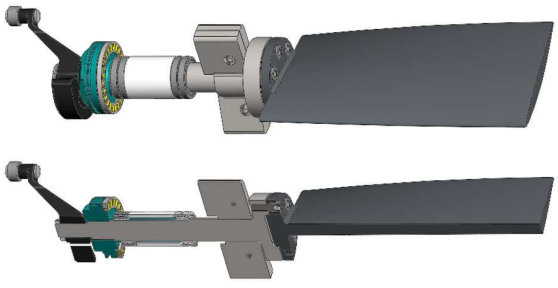


Figure 6: The blade.

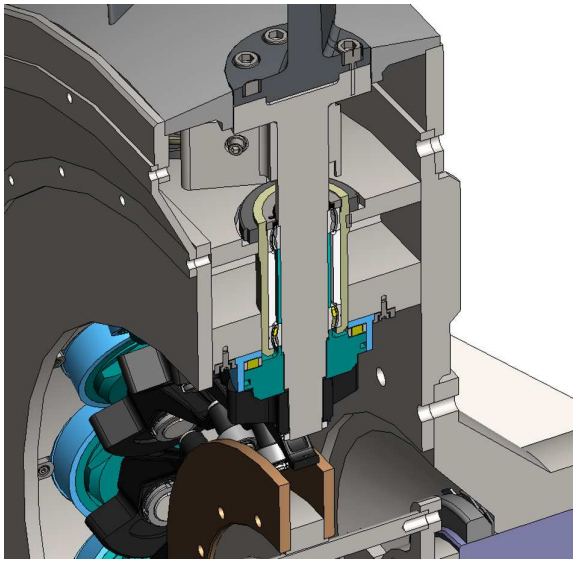


Figure 7: Blade assembly axis.

shaft a counterweight, made by two shells, and two radial spherical roller bearings are assembled. At the bottom of the shaft, an axial roller bearing and its supports are screwed to the shaft itself. At the end of the shaft the crank is assembled. The blade shaft, without the crank and the support of the axial bearing, is inserted in a cylindrical element, in which the radial bearing housings are obtained. This solution makes the assembly of the blade independent from the assembly of the impeller. Moreover it is possible to balance each blade-shaft individually and do maintenance in a simple way. The position of the cylindrical element is defined by a mobile mechanical stop which

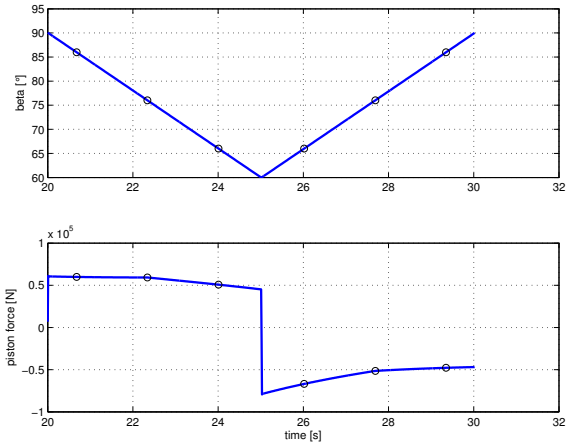


Figure 8: Piston force.

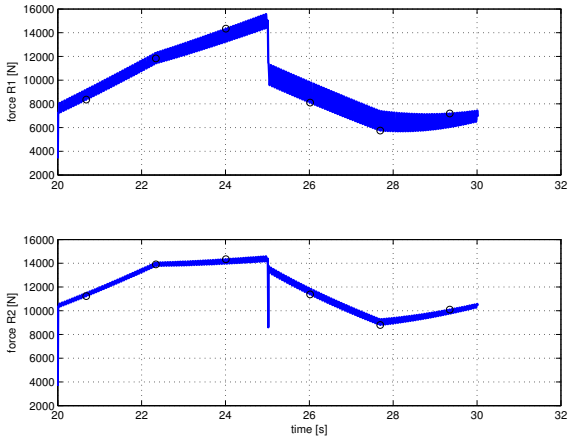


Figure 9: Absolute value of the reaction force R_1 and R_2 (internal and external bearing of the blade shaft).

is made by a threated ring which preload the axial bearing as shown in figure 7. It can be also observed that the lubrication system is realized thanks to a single grease nipple. The blade is made of aluminum and it is 640 [mm] high; between the base and the tip there is a twisting of 20◦. The weight of the blade is 16 [kg], while the blade and the shaft with the radial bearings housings weighs 62 [kg]. Figure 10 shows the full-assembled fan. The stem of the oil-hydraulic actuator is rigidly linked to the impeller, while the sleeve is mobile. On the sleeve two plate are mounted: they have to guide the wheel on the tip of each crank.

5. MULTIBODY MODEL AND RESULTS

To evaluate the loads on the bearings on the most stressed mechanical parts and to estimate the force that the actuator must be able to supply, a multibody model of the fan has been created with the simulation environment Matlab-SimMechanics, in which the aerodynamic forces were included. These are available [1] for some areas of operation and for some values of the blade rotation. Intermediate values were obtained by interpolation.The friction model due to the bearings have been obtained applying a torque in opposition to the blade rotation in correspondence of the axial bearing. Friction torque is calculated with relation and coefficients of the chosen bearings manufacturer: $C_{\text{att}} = K \cdot f_{\text{SKF}} \cdot r_m \cdot F_a$. Where $f_{\text{SKF}} = 0.005$ is the friction coefficient obtained from the catalogue; r_m is the internal radius of the axial bearing of 60 [mm]; F_a is the force in the direction of the blade-shaft axis. It can be observed that in the previous relation a safety coefficient has

[1] The blade profile has been obtained from the dynamic optimization on the base of the fan requirements

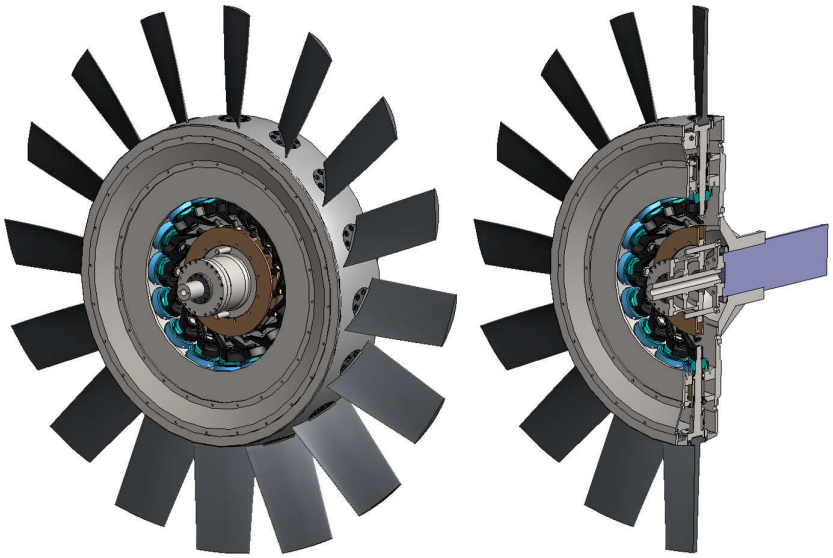


Figure 10: Complete draw of the fan.

been introduced ($K = 3$) to consider friction phenomena due to assembly inaccuracies or due to unconsidered deformations or friction (such as frictions between the crank and its wheel or between the wheel and the plate).

Some results obtained by the multibody model are reported in this section. In particular in the simulations the impeller starts from a velocity of 0 [rpm] and in 10 seconds it reaches 1000 [rpm]. In this phase the blades are kept closed, until the operating speed is reached, in this way aerodynamic forces can be neglected and the reaction forces on the impeller bearings are constant.

When the rotor has reached the operating velocity, the blades are opened. For example we report a regulation phase around the operating point (figure 8), which corresponds to the orientation of the crank β of $76°$. In the range of $\pm 30°$ it is possible to interpolate the values of the known aerodynamic forces. Since the regulation of the angle of incidence of blade is very slow, the inertial torque which is generated by the rotation of the blade among its own axis is negligible if compared to the torque due to the aerodynamic effects and the friction torque.

The graphics in figures 8 and 9 show respectively the force provided by the actuator and the loads on the bearing of the blade-shaft in the radial direction. The axial load is about of 560000 [N] on the inner bearings mostly due to the centripetal force. Thank to the results obtained from the simulation with the multibody model, it was possible to calculate correctly the size of the bearings and evaluate the loads on the impeller to define the size of the structural parts.

6. CONCLUSION

This article presented the design process and the computation tools that are useful for the project of a variable pitch axial fan. Particular importance was given to the kinematic synthesis and the dynamic analysis of the command mechanism of the rotation of the blades. the implementation of a multi-body model allowed us to simulate the inertial effects, friction and aerodynamic loads to assess the most critical components and appropriately size them.

REFERENCES

[1] Th Billotet and U. Johaenntgen. Bexbach ii power station - advanced electricity generation from bituminous coal. *VGB-Kraftwerkstechnik*, 75(1):21–25, 1995.

[2] Abroell G. Electricity from pit coal - bexbach power plant. *BBC-Nachrichten (Brown, Boveri and Cie)*, 66(4):141–148, 1984.

[3] Gurbin and Florian. Adjustable pitch fan. *Patent, Canada*, 1978.

[4] W.B. Kolb. Lubrication of fan blade bearings. *Patent, Germany*, 1974.

[5] L. Mulluer. Variable-pitch axial flow fans for thermal power station.

[6] A. Takata. Construction of adjustable blade shaft bearing in axial-flow fan with adjustable blades. *Patent, Japan*, 1979.

Multidisciplinary design methodology: an intermediate representation tool based on virtual reality

Gaël Guerlesquin
Laboratoire Systèmes et Transports
Université de Technologie de Belfort-Montbéliard
Montbéliard, France

Morad Mahdjoub
Laboratoire Systèmes et Transports
Université de Technologie de Belfort-Montbéliard
Montbéliard, France

and

Jean-Claude Sagot
Laboratoire Systèmes et Transports
Université de Technologie de Belfort-Montbéliard
Montbéliard, France

ABSTRACT

Nowadays firms have to use concurrent design to meet customers' expectations. A plural approach to the design process is essential. Unfortunately, the design activity involves experts specialized in various aspects of the product such as aesthetics, ergonomics and mechanics. Thus the collaboration between these experts is particularly difficult in the convergence steps and often results in complex adjustments.

This paper details a proposition of design methodology based on a multidisciplinary approach, using virtual reality tools. Our goal is to facilitate the integration of ergonomics and aesthetics in a mechanical design process. In this context, we consider virtual reality as an intermediary design tool useful for collaborative decision support during convergence phases. We present our methodology and associated tools tested during an industrial project, by focusing on an aesthetics-mechanics convergence step.

Keywords: Design methodology, virtual reality simulation, user integration, mechanical design.

1. INTRODUCTION

In recent years, manufacturing companies have been innovating in order to improve competitiveness and business performance. They must introduce innovative products into the market more efficiently and faster to maximize customer interest and sales. Moreover, engineers cannot always treat qualitative data they have at their disposal such as ergonomic and aesthetic characteristics. Thus, it is often through common sense rules that designers integrate these characteristics coming from different needs. As a consequence, many products surrounding us have not been designed to respond to end-user needs [1]. Many needs exist, affecting future product functionalities (ergonomic, quality, aesthetic, commercial, etc.). Each one is studied by specialists. In our study, we focus on ergonomic and aesthetic needs. Ergonomic needs correspond to the expression of the user's demand, they match the product's usefulness and they are related to all usage notions, according to the physiological and psychological characteristics of the user [2]. The aesthetic needs are linked to the user's expectations. They are often considered secondary needs because they are not essential to the product usability or performance. Nevertheless, they allow the product to be better adapted to the customer and therefore sell better [3]. Jordan calls them the "emotional benefits" [4].

Ergonomics and aesthetics require a specific integration in design process because of large culture differences between specialties. Thus, the product technical designers require the help of specialists to treat aesthetic and ergonomic needs. Unfortunately, there are few methodologies allowing them to work together and it is very difficult for designers to combine ergonomic and aesthetic features.

This paper depicts the first results of a research work on the articulation of design disciplines, and in particular ergonomic design and aesthetic design. We therefore present, in a first part, a new design methodology. This methodology is based on a proposal for a global approach to the design process, including the three actors we consider: the ergonomic designer, in charge of the ergonomic needs and the user characteristics, the aesthetic designer, in charge of the aesthetic needs, and the mechanical designer, responsible for the technical aspects of the product and leading the multidisciplinary convergences. In a second part, we will present a case study emphasizing the virtual reality technology benefits in this methodology and more particularly in the convergence phases. In this first experiment, we use a new design tool based on virtual reality technologies to support our methodology. Preliminary results show that this new approach paves the way for a new type of collaborative design.

2. TOWARDS A MULTIDISCIPLINARY DESIGN METHODOLOGY

Aesthetics and ergonomics integration in design process

For many years, design methodology was oriented towards the technical aspects of products. The aim was to increase profits by optimizing the process in accordance with the quality-cost-delay triad which left out the final-user needs and wants [2]. Today, mechanical designers, specialists of these technical aspects, need to be helped by specialists of ergonomic design and aesthetic design, considered as co-designers [5]. Each of them has their own specific tools and methods used at different steps

of the design process [6] [7] [8]. Unfortunately, the collaboration is complex and the tools are generally not compatible with one another because of the culture and vocabulary differences [9]. Moreover, designers do not rely on the same data type: ergonomic designers use some numbered prescriptive criteria necessary to mechanical designers, whereas aesthetic designers apply subjective ratio and equilibrium. However, a collaborative work is essential to design a product that satisfies all specifications (mechanical, ergonomic and aesthetic) [10]. Thus, it is necessary for designer teams to set a methodology based on the differences between various approaches and tools. In order to develop products in accordance with the various approaches mentioned, we base our work on DFX methods. DFX (Design For X) is a succession of product development techniques which can be applied effectively to the design process. It allows not only the rationalization of the products, but also of the associated processes and systems [11]. In this work, we will focus more particularly on Design For Usability (DFU) [12] or Design For Ergonomics (DFE) [13], and Design For Pleasure (DFP) [4] or Design For Aesthetics (DFA).

Multidisciplinary design methodology

In order to allow a better cooperation of experts during the design process, we propose a multidisciplinary global design methodology (Fig.1). This methodology has already been implemented in seven urban equipment designing projects to be tested [14]. Our objective is to integrate ergonomics and aesthetics, which are essential in a human-centred design process. Thus, the ergonomic approach based on ergonomic needs, and the aesthetic approach based on aesthetic needs, are organized around a central support, the mechanical design. The design methodology proposed includes the multidisciplinary interactions with the aim to anticipate and develop a set of specific tools for each step of the design process. These interactions are the subject of reasoned discussions and compromises often difficult to assess. They occur generally in convergence steps, signalled by double arrows in fig.1 [8].

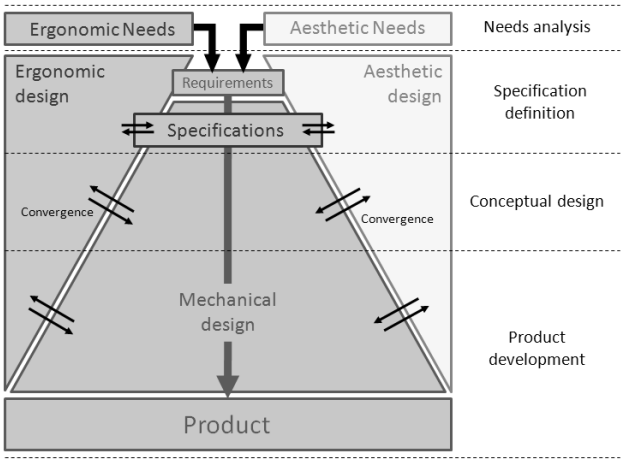


Figure 1: Multidisciplinary global design methodology [14]

Several experiments have shown that our methodology required convergence tools for designers to improve the collaborative aspect. In this manner, it is important to develop new collaborative tools in order to allow a better multidisciplinary convergence, accommodating the importance of expertise. The intermediate representations appear to meet this requirement [15].

Virtual reality as intermediary tool for help multidisciplinary convergence

The concept of intermediary representation exists in numerous design models and allows the transition between the design process steps as well as communication between experts [15]. According to Pascal [16] the intermediary representations fulfil three functions: i) the object **representation** within its environment and its background as well as its end use, ii) the **translation** giving a universal or multidisciplinary understanding, iii) the **mediation** because these objects are necessary to take a decision. Thus, drawing, cardboard mock-up, prototype, listing and screen copy are considered as intermediate representations useful in the design process. Each expert has their own intermediate representations enhancing various aspects of the product.

In order to represent and test the product, we operate in the virtual prototyping area [17]. The numerical aspect is a guarantee of reactivity and simplifies the object evolution through the various steps of the design process. However, the tools for conventional virtual representation such as 3D-CAD or 2D-realistic are not elaborated for a collaborative use. They do not allow optimal communication between experts. In this context, a universal numerical representation or a more flexible representation would provide each partner with an adequate tool for collaborative design.

In our work, in accordance with numerous authors [18] [19], we propose to use virtual reality tools in the design process and more particularly in the convergence phases. We assume that the collaborative aspect of immersive virtual reality facilitates the **mediation** and the **translation**, and allows all experts to simultaneously interact on a same model. Indeed, Virtual Reality adds to CAD-model properties, some VR specificities, such as immersion and interaction in real-time. Some authors have proposed virtual reality tools to try to address the convergence problem [20], but without any real multidisciplinary design methodology. For example, Krause proposes a system for creating and modifying the virtual prototype in real-time [21]. These applications are generally developed with the aim of helping aesthetic designers to formalize proposals. Unfortunately, it remains difficult for mechanical designers to integrate these proposals in their own work.

3. CASE OF STUDY: AESTHETICS-MECHANICS CONVERGENCE WITH VIRTUAL REALITY TOOLS

Context of study

In order to test our methodology and VR associated tools, we applied it to a multidisciplinary project involving the design of a new product: the F-city vehicle. F-city is a small urban car constructed by FAM, a French company specialized in automotive subcontracting. It is equipped with two seats and an electric motor. The design of this new concept car required a multidisciplinary approach and we therefore deployed our methodology. The design team was composed of five mechanical designers, two ergonomic designers and three aesthetic designers.

The case study presented deals with an aesthetics-mechanics convergence: validating materials configuration, according to colour and texture propositions. This convergence arose during the product development phase (Fig. 1). Study partners were

one aesthetic designer and one mechanical designer. The elements involved were the vehicle body and interior equipment: dashboard, seats, inner liners, etc. Virtual Reality was used as a collaborative intermediary representation tool and an interactive and reactive design system. In the related convergence phase, we focused on materials validation.

Procedure

First, the aesthetic designer analyzes aesthetic needs and makes suggestions (Fig2.A). To facilitate the information transmission to the mechanical designer, the aesthetic designer has a VR application for texture and colour configuration. The application allows configuring vehicle elements while enjoying the VR benefits in terms of scale and a real-time simulation. Once suggestions are formalized, colour data are extracted in TXT format (CMYB). Textures, predefined in a specific library, are also listed.

The extracted data are integrated directly into the CAD modeller and the mechanical designer can infer acceptable materials. For every possible material, the mechanical designer tests the technical behaviours to validate the best choice (Fig2.B). When materials are defined, the mechanical designer dumps his CAD data (CATIA Model) to VR software (VIRTOOLS). Corresponding colours and textures are exported toward the application. The received file can be imported into the application, putting preconfigured colours and textures on the selected vehicle part.

From then on the technical proposals are submitted, a multidisciplinary convergence phase is carried out in virtual reality (Fig2.C). Once more, the application for texture and colour configuration is used. An additional mode dedicated to materials allows the mechanical designer to simulate real-time display of the various technical proposals. The tool is used as a communication tool, allowing, in our opinion, the aesthetic designer and the mechanical designer to reach together the best solution integrating the aesthetic and technical characteristics.

When a satisfying proposal has been reached, texture/colour/material characteristics are exported toward a technical specification board.

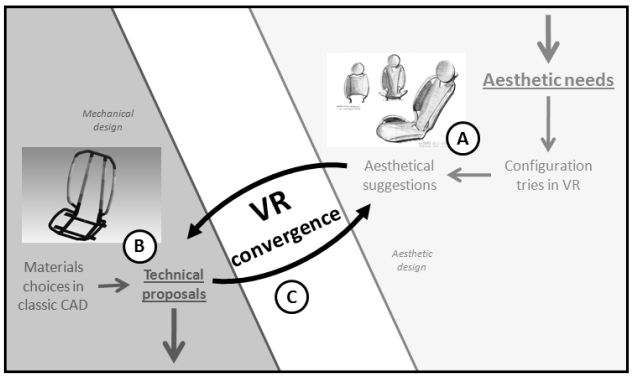


Figure 2: Virtual Reality Aided Multidisciplinary Convergence method

Equipment used

As shown in fig.3, during our experiment, we use a virtual reality platform composed of 3 active stereoscopic screens (2.10m * 2.80m). Images are displayed at 115 Hz with a 1400 x 1050 pixel resolution with Christie Mirage S + 4K projectors. The stereoscopic images are generated by 3 graphic computers equipped with Nvidia Quadro FX 5600. The motion tracking is realized by an ART optical system which is composed of 6 cameras and numerous trackers for user body. The hand gesture is recorded / tracked thanks to 5DT data gloves and for more precise gestures we can use ART Finger tracker.

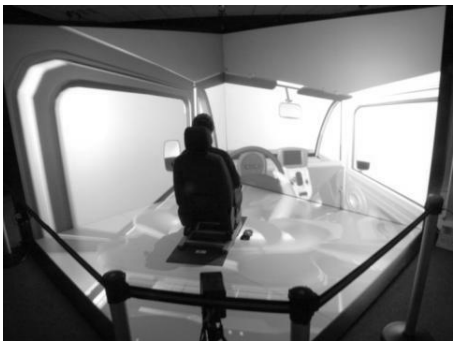


Figure 3: Virtual reality platform used in our approach

A tactile tablet of PC type was also used to support the colour-texture-material application.

Developed application

The used application was implemented following an interface design methodology [22]. The principle was to provide two interfaces on the same 3D model, one for the aesthetic designer, and the other for the mechanical designer. The application allows the aesthetic designer to control the texture and colour of the vehicle body, dashboard, seats, opening, with the interface shown in fig.4. The designer interacts with the virtual model using an interface integrated in the tactile tablet PC.

Thus, thanks to the application, the aesthetic designer can move himself or the 3D model in several virtual environments (Fig 4A). He can modulate the colours from Cyan, Magenta, Yellow and Black (CMYB), and see the results in real time, with the exterior (Fig 4E) independent from the interior (4F). Moreover, the application allows obtaining a simulation of the lights and the opening movements (Fig 4B). Finally, it is possible to export the chosen configurations using a global backup (Fig 4C).

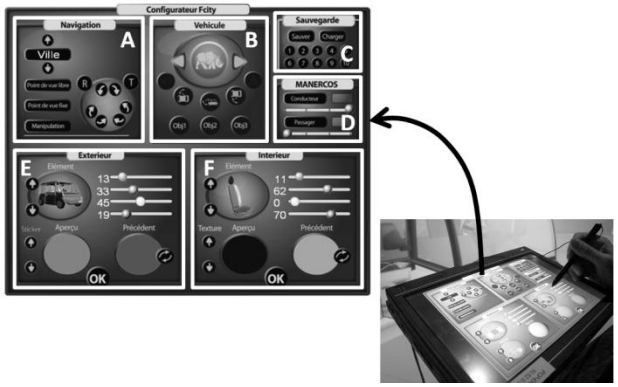


Figure 4: Configuration colour-texture interface

The mechanical designer has a specific mode at his disposal, allowing the display of technical proposals. The selection of one technical proposal sets the terms of colour characteristics.

Experiment results

In agreement with the instructions for the marketing positioning required by the company, the aesthetic designer first worked on the product trend and on the product environment to understand the aesthetic needs of the customers. This work led him to propose various graphic universes, represented by trend boards and product boards.

In the first step of the experiment, the aesthetic designer made configuration choices on the virtual reality platform, according to the trend proposals. He had his trend boards and the tactile tablet to help him to make proposals (Fig.5).

To illustrate this paper, following are three examples: body sides, seats and window posts. In every case, the aesthetic designer tested different colour and texture configurations and saved them when they were satisfactory. The CMYB manipulation was a request from the aesthetic designer.

Figure 5: Texture-colour configuration in Virtual Reality

After a creativity phase, the aesthetic designer defined colours and textures characteristics for each considered part. Those characteristics are shown on the Table.1.

Vehicle part	Colour	C	M	Y	B	Texture
Body sides	Cream	0	0	10	5	Matt and smooth
Seats	Brown	20	35	50	0	Large stitch
Window posts	Grey	8	6	6	0	Shiny and smooth

Table 1: Colours and textures selected by the aesthetic designer

The second step of the experiment was the data transmission toward the mechanical designer. He based the materials choice on the aesthetic designer suggestions, according to the technical requirements. He has selected painted steel for body sides to respect the cream colour requested. The seat has been equipped by a brown fabric, to respect the aesthetic designer wants. As regards window posts, they are structural parts requiring high resiliency. Thus, the mechanical designer gave priority to the material yield, selecting 22MnB5 Steel, darker than the requested colour. Selected materials and associated colours are showed on the Table 2.

Vehicle part	Material	Surface finishing	C	M	Y	B
Body sides	Steel	Painted : White 9010	0	0	10,2	5,1
Seats	Fabric	Sattler 804 K73	18,8	36,9	56,2	0
Window posts	Steel 22MnB5		58,1	43,5	41,9	7,1

Table 2: Material and surface finishing suggested by the mechanical designer

Once the mechanical designer had defined the materials based on the aesthetic designer's suggestions, a convergence/validation phase was required (Fig. 2C). The two designers used a VR platform to compare their proposals and assess the gaps to reach a solution. The interface was handled by them both and offered proposals which allowed for further changes. Furthermore, the aesthetic designer used his trend boards, while the mechanical designer used a selected reference materials book, and colour variants for each one.

Initially, the evaluation began with a quick assessment by the aesthetic designer. He appropriated the mechanical designer's proposals, by manipulating the object and moving some parts of the model.

In the next phase, the aesthetic designer compared the materials set proposals with a trend board, and then examined the aesthetic proposals, to assess differences. Some comparison between proposals was made possible thanks to VR (Fig.6).

Figure 6: Seat colour comparison

Finally, the differences observed were discussed. As regards body sides, the colour suggested by the aesthetic designer could be adopted thanks to a large measure of freedom existing in the paint ranges. Unfortunately, the matt texture could not be accepted because it was incompatible with the smooth appearance requested by the aesthetic designer. However, during the discussion, the mechanical designer pointed out that this choice was arbitrary, and could be changed. Thus, following the mechanical designer's recommendations, the aesthetic designer chose an epoxy-polyester powder coating which allowed obtaining a sanded matt texture.

As regards window posts, there was a convergence problem because the aesthetic designer's desired colour could not be adopted. The technical choice being unsatisfactory, it became subject to a reasoned discussion. The objective of the aesthetic designer was to find another material that offered similar yield. Unfortunately, the mechanical designer had already studied various materials and concluded that no flexibility was possible on the window post, the driver's safety being a priority. For this reason, the aesthetic designer allowed his suggestion to be rejected and the mechanical designer's proposal was chosen.

Once a satisfactory configuration was reached, the two designers recorded the proposal. The exported data were the characteristic CMYB numbers of each component and some views of the different states. A report on the selected materials was also produced. Fig.7 shows one of the final configurations, which has been manufactured.

Figure 7: Final selected configuration for F-City manufacturing

Discussions

The textures and colours configuration is a representative step of aesthetics-mechanics convergences because it involves accommodations with the materials and surface treatments. The application developed has allowed us to highlight the interest of equipping designers with convergence tools based on virtual reality. Convergence phases are always difficult because of the confrontation of proposals coming from various experts. Our new application allows designers to reach compromises to satisfy aesthetic needs, according to technical requirements. The virtual reality platform permits each partner having a product representation according to his usual design methods. It also allows everyone to test the product according to his own criteria for technical solutions and to evaluate and validate the final product.

The basic model (Fig.1) depicts the articulation of the design disciplines involved. It brings out multidisciplinary informal interactions and this without specific tools. Thanks to this first experiment, we can introduce a new tool based on virtual reality. It allows formalizing exchanges between the different designers. Thus, new interactions appear which result in a new way for the mechanical designer to integrate the expertise of the aesthetic designer in the design process.

4. CONCLUSION AND PERSPECTIVES

The F-City project was a first test for our global design methodology, positioning Virtual Reality as a collaborative intermediary representation. Our first results demonstrate that the virtual reality platform can offer a narrow collaboration between the designers, providing a fast and pertinent convergence for design. In this approach an important point is that the aesthetic designer uses his expertise in an immersive virtual reality environment. Scale 1 had a strong effect on the perception of the proposed concepts and their evaluation.

The originality of our study is the importance of virtual reality in aesthetics-mechanics convergence. Our approach is thus a new way to combine a global design methodology and a new collaborative tool for intermediary representation. Another aspect which we consider to be important is the advantage of being able to extract the qualitative data coming from the aesthetic designer, in the convergence step. In addition to the integration of each expertise, this approach allows the mechanical designer to better interpret and appropriate the qualitative data coming from aesthetics analysis.

Nevertheless we have been able to target several improvements to be brought to the proposed application. Those improvements concern in particular the integration of functionalities dedicated to mechanical designers. The configuration tool could allow the integration of company manufacturing capacities. Thus, we propose to predefine characteristics of standard materials and available surface treatments, and to associate automatically textures and colours. Finally, we propose some measure of freedom indication for the material choice when colour and texture are chosen.

5. ACKNOWLEDGEMENTS

The authors acknowledge Florence Bazzaro, Assistant Professor in Cognitive Sciences and Virtual Reality, and Lucien Seichepine, computer graphics technician from SeT-ERCOS, for developing a set of applications for this work and for implementing them in our experiments. Maxime Larique and Jérôme Maysse, mechanical design engineers, Marjorie Charrier and Hugues Baume, researchers in industrial design, helped us with the F-City project and the convergence studies. Finally, we thank FAM Company for permitting us to experiment our methodology on the F-City project.

REFERENCES

[1] Norman, D.A. *Emotional Design, why we love (or hate) everyday things*. Ed Basic Books. 2004. ISBN: 0-465-05135-9.

[2] Norman, D.A. *The design of everyday things*. Ed Basic Books. 1988. ISBN: 0-465-06710-7.

[3] Jordan, P.W. Human factors for pleasure in product use. *Applied Ergonomics*. 1998. Vol. 29, No. 1, pp 25-33.

[4] Jordan, P.W. *Designing pleasurable products, an introduction to the new human factors*. Taylor & Francis. 2000. ISBN: 0-415-29887-3.

[5] Duchamp, R. *Méthode de conception de produits nouveaux*. Paris: Hermès Science Publications. 1999. ISBN: 2-7462-0045-7.

[6] Quarante, D. *Eléments de design industriel*. Paris: Polytechnica. 2001. ISBN: 2-84054-018-5

[7] Sagot, J.C., Gouin, V., & Gomes, S. Ergonomics in product design: safety factor, *Safety Science*, 2003. 41(2-3), 137-154.

[8] Ullman, D.G. *The Mechanical Design Process*. McGraw-Hill Higher Education. 2003. ISBN: 0-07-237338-5.

[9] Kvan, T. Collaborative design: what is it? *Automation in Construction* 9 (2000) 409-415.

[10] Pahl, G. & Beitz W. *Engineering Design: a systematic approach*, K. Wallace ed., Springer 88, 2nd edition 1996. ISBN: 1-8462-8318-3.

[11] Huang G.Q. & Mak K.L. The DFX shell: A generic framework for developing design for X tools. *Robotics & Computer-Integrated Manufacturing* 1997; 13(3):27180.

[12] Nieminen, M. *Information support for user-oriented development organization. Considerations based on the construction and evaluation of knowledge storage*. Espoo (Finland): Helsinki University of Technology. Thesis. 2004.

[13] Allada, V., Kopardekar, P., Anand, S. & Mital, A. Implications of ergonomic and user considerations on manufacturing consumer. In: DE flexible assembly systems, *proceedings of the fourth international conference on design, theory and methodology*. 1992. pp 25-33.

[14] Guerlesquin G. & Sagot, J.C. Vers une meilleure intégration de l'ergonomie et du design industriel dans la conception. *6ème Conférence Internationale Conception et Production Intégrées*. Fès (Morocco), October 2009.

[15] Jeantet, A. & Vinck, D. "Mediating and commissioning objects in the sociotechnical process of product design: a conceptual approach", pp. 111-129, in MacLean, D., Saviotti, P., Vinck, D., *Management and new technology!: design, networks and strategies*, Cost A3, Vol.2, Bruxelles 1995.

[16] Pascal, A. & Thomas, C. Appropriation des TIC: vers une méthodologie de coconception orientée usage: le cas KMP. *Ouvrage collectif post-journée, Journée de Recherche: L'appropriation des outils de gestion: vers de nouvelles perspectives*, Presses de l'Université de Saint Etienne. 2006.

[17] Wang, G.G. Definition and review of virtual prototyping. *Journal of Computing and Information Science in Engineering*, September 2002. Vol. 2(3), pp. 232-236.

[18] A. Kadri, H. Christofol, & H. Samier, Virtual Reality Application for Collaborative Design Evaluation: Virtual Dustbin Design, *Proceedings of International Conference on Virtual Reality (VRIC)*, Laval, 2004. pp. 119-127,

[19] Mahdjoub, M., Gomes, S., Sagot, J.C. & Bluntzer, J.B. Virtual Reality for a human-centered design methodology. *6th Eurosim (federation of European simulation societies) congress on modeling and simulation*. Ljubljana, Slovenia. September 9 - 13, 2007.

[20] Bordegoni, M., Colomboa, G. & Formentinia, L. Haptic technologies for the conceptual and validation phases of product Design. *Computers & Graphics*. 30 (2006) 377–390.

[21] Krause. A Three-stage Conceptual Design Process Using Virtual Environments. *SCG'2004*, February 2-6, 2004, Plzen, Czech Republic. (Technische Universität Berlin)

[22] Bennes, L. & Bazzaro, F. & Sagot, J.C. Multifunctional Virtual Screen: a 3D User Interface for stylist and mechanical engineer. *Actes des cinquièmes journées de l'Association Française de Réalité Virtuelle*. 2010.

Mathematical Optimization in Multidisciplinary Design Through A New Decomposition Approach: A Case Study of an Aircraft Design

Nand K. Jha
Professor
Mechanical Engineering Department
Manhattan College, Riverdale, New York 10471

Abstract

A new mathematically rigorous decomposition approach is presented for optimization in multidisciplinary design. The approach suggests identification of coupling and behavioral variables in the subsystems of the overall optimal design. The proposal presents a methodology for forming a coupling subsystem which estimates the coupling variables and then passed on to different subsystems of aerodynamics, propulsion, and structural subsystem. There are some behavior variables in the multidisciplinary design problem of aircraft and it is proposed that they be transferred (communicated) between relevant subsystems and that would make it easier to analyze the multidisciplinary design optimization (MDO) problems. The coupling and behavior variables identification should make the analysis of aircraft multidisciplinary design more adaptable. The concept of automatic mathematical modeling for MDO has also been introduced to automate the entire multidisciplinary design process. The approach is based on geometric programming technique and suggests reduction and elimination of loose and redundant constraints. After solving the subsystems, finally the overall objective function value is obtained. The MDO formulation and solution presented in the proposal is based on new methodology and concept.

Keywords: Multidisciplinary Design Optimization (MDO), Geometric Programming, Decomposition Approach and Aerodynamics

1. Introduction

Aircraft design is multidisciplinary, that is, it requires the coordination of information from number of highly specialized disciplines. It may include the disciplines of aerodynamics, structures, propulsion, controls, and manufacturing. The design approach or emphasis of each discipline specialist may be quite different. As design problem becomes more complex, the role of disciplinary specialists increases and becomes more difficult for a central group to manage the process. As the analysis and design tasks become more decentralized, communications requirements become more severe. These difficulties with multidisciplinary design are particularly evident in design of aircraft.Since each system in an engineering design consists of multiple disciplines linked (coupled) together a design process should consider all of these systems concurrently. In addition to optimize the design given the resources and time at hand, design optimization should take into account all relevant disciplines simultaneously. Multidisciplinary Design Optimization (MDO), which simultaneously takes into account mutually dependent design elements from various fields, has the advantages of reduced time and cost compared to serial approaches. MDO was originally devised and developed for designing aircrafts, though it has recently been applied to diverse design applications, including ship building and automotive engineering. Applying MDO methodologies to actual design problems requires the organization of various analysis tools and other resources, optimization aids, and Computer Aided Design (CAD) software. Considerable time is required for the appropriate integration of these elements.

Therefore, an integrated MDO environment is necessary to conduct such processes more conveniently and efficiently. Designing advanced engineering systems like aircrafts, is fundamentally a complicated process; it consists in essence out of a lot of involved and interwoven elements. This Multidisciplinary Design Optimization (MDO) of such large systems is characterized by the interdisciplinary couplings, multiple objectives, a large design variables space and a significant number of design constraints.

Kroo [19] describes the variants of collaborative optimization being applied to large-scale problems. The strategy of distributed design optimization is also very appealing. However, sequential disciplinary designs and informal iterations can lead to sub-optimal design. It is for this reason the methodology presented in this proposal includes not only the coupling variables but also behavioral variables which move back and forth between different subsystems.

Sobieski et al [20] presentation of bi-level integrated system synthesis (BLISS) is a definite improvement in this direction due to bi-level integration concept. However, it still comprises of large level local design variables and the affect of other subsystems on this particular solution is not analyzed. The interactions between different subsystem and its interrelation seem to be lost. MDO is the interaction of design engineering, material science, control systems, thermal science, and a robust optimization technique. For such a complicated problem, there is need for development of a mathematically rigorous algorithm and interaction with finite element analysis. The algorithm should be generic so that any kind of aircraft design problem could be dealt with. Significantly design should produce a part of high quality at minimum cost. The approach should also be flexible and easy to use. It may be prudent to develop comprehensive model (i.e. a model encompassing all important objectives outlined in Airforce Research Proposal). The new procedure developed should represent improvement over the old ones i.e, it should be adaptive. However, it should be emphasized that the model proposed in this proposal is conceptual only.

The method proposed in this project is based on geometric programming technique [1-3]. The system geometric program is solved through a decomposition approach as presented below. The solution procedure proposed here through the decomposition approach may require the solution of a large number of geometric programs or to be exact almost as many as subsystems.

The advantage of a solution procedure such as proposed here is that it takes into account the interaction and interrelations of different subsystems and their combined effect on the system. The following geometric program is proposed:

2. Mathematical Formulation and Algorithm

Convention Adopted in Writing the Equations in MDO:

Coefficients:

$C_{mt}^{\ l}$: l = 0, 1,2,3,4 (refers to subsystems)

$m = 0,1,2,.....M$ (Refers to the equation numbers starting with 0, the objective function and constraints start with number 1).

a_{mtn} : Exponents of the variables, n

t is the number of terms in an equation.

So a_{mtn} refers to the exponent of n th variable in equation m of term t.

T_m refers to total number of terms.

General Formulation of the MDO: It is required to

Maximize

$$\sum p_r \left(\sum_{t_1}^{T_{000}} C_{0t}^{\ r} \prod_{n=1}^{h} (X_n^{'})^{a_{0tn}} \prod_{n=1}^{N=h} (X_n^{"}(r))^{b_{0tn}} \right) \tag{1}$$

Subject to

$$\sum_{m=1}^{M} C_{mt}^{\ r} \prod_{n=1}^{h} (X_n^{'})^{a_{mtn}} \prod_{n=1}^{N-h} (X_n^{"}(r))^{a_{mtn}} \le 1 \tag{2}$$

Where R is the possible number of subsystems with coefficients $C_{mt}^{\ 1},.....,C_{mt}^{\ R}$ vectors with known weightage $p_1,....,p_R$ such that

$$\sum_{r=1}^{R} p_r = 1, \qquad p_r > 0, \tag{3}$$

$X_n^{'}$ are the design variables, which are fixed through the operation of the system or independent of the subsystems.

The determination of the design variables as well as operating variables in different subsystems may be likened to the minimization of a system objective function under the subsystem constraints. The design variable vectors ($X^{'}{}_n$) in the system corresponds to the coupling vectors operating variables under different subsystems. With this concept, in view of the other variables in the subsystem program is converted to the following standard:

Design variable (two stage program) ≈ coupling vector ($X_n^{'}$)

Operating variable in first subsystem ≈ 1st subsystem vector ($X_n^{"}(1)$)

Operating variable in second subsystem ≈ 2nd subsystem vector ($X^{"}{}_n(2)$)

Operating variable in 3rd subsystem ≈ 3rd subsystem vector ($Xn''(3)$)

……

Operating variable in Rth subsystem ≈ Rth subsystem vector ($X_n^{"}(R)$).

Each subsystem function is independent except for a coupling vector ($X_n^{'}$), which is common to all subsystems with a fixed value of ($X_n^{'}$), hence, the total system decomposes into ℓ independent subsystems. Each subsystem ℓ =0, 1, 2, R, may have a system function and constraints which depend on the vector ($X_1^{'}$) and $X_n^{"}(\ell)$. Similarly, a coupling subsystem, denoted by an ℓ =0, contains the coupling vector ($X_n^{'}$) only is defined, which also may have an objective function and a set of constraints. The objective function and constraints of all the subsystems are required to be of the form as follows;

Let $C_0^{\ \ell}(X_n^{'}, X_n^{"}(\ell)),$ be the objective function associated with ℓ th subsystem, ℓ =1, 2, R, be given by

$$C_0^{\ \ell}(X_n^{'}, X_n^{"}(\ell)) = \sum_{1}^{000} C_{0t} \prod_{n=1}^{h} (X_n^{'})^{a_{0tn}} \prod_{n=1}^{N=h} (X_n^{"}(r))^{a_{mtn}} \tag{4}$$

The constraints associated with the ℓ th subsystem are expressed as

$$C_{mt}^{\ \ell}(X_n^{'}, X_n^{"}(\ell)) = \sum_{1}^{000} C_{mt}^{\ \ell} \prod_{n=1}^{h} (X_n^{'})^{a_{0tn}} \prod_{n=1}^{N=h} (X_n^{"}(\ell))^{a_{mtn}} \le 1 \tag{5}$$

$$X_n^{'} > 0, X_n''(\ell) > 0$$

Similarly the objective function of the coupling subsystem is written as

Maximize

$$C_0^{\ 0}(X_n^{'}) = \sum_{1}^{000} C_{0t}^{\ 0} \prod_{n=1}^{hN} (X_n^{'})^{a_{0tn}} \tag{6}$$

Under the constraints

$$C_k^{\ 0}(X_n^{'}) = \sum_{1}^{000} C_{mt}^{\ 0} \prod_{n=1}^{N} (X_n^{'})^{a_{mtn}} \tag{7}$$

The system optimization problem is then to calculate values of $X_n^{'}$, $X_n^{"}(1),.....,X_n^{"}(R)$ which maximizes the total

system (all-in-one) objective function under the real world constraints given by

$$C = C_0^{\,0}(X_n^{'}) + C_0^{\,1}(X_n^{'}, X_n^{''}(1)) + ...C_0^{\,2}(X_n^{'}, X_n^{''}(2)) + ... + C_0^{\,R}(X_n^{'}, X_n^{''}(R)) \quad (8)$$

Under the constraints

$$C_k^{\,0}(X_n^{'}) \le 1, \quad (9)$$

$$C_k^{\,\ell}(X_n^{'}, X_n^{''}(\ell)) \le 1, \ell = 1, ..., R \quad (10)$$

The systems optimization problem is a primal geometric program with a special structure. It is assumed that the problem is well defined and formulated for which the experts working on the topic are the best source.

The objective is to solve this system's optimization problem by decomposing it into subsystems. The algorithm followed for system's optimization in terms of subsystem optimization is as follows.

(1) Obtain the subsystem's optimization through any standard geometric programming technique. In this work, an augmented geometric program approach is followed for the subsystem's optimization. After optimizing different subsystems, m_l, the minimum value of ℓ th subsystem and coupling subsystem are obtained.

(2) M = the maximum value of the complete (all-in-one) system

$$= \left(\sum_{l=1}^{R} (m^{\ell})^{1/\tau} \right)^{\tau} \quad (11)$$

Where $\tau = 1 - \sum \omega_j^{\,*}$ and $\omega_j^{\,*}$ are optimum dual variables associated with coupling subsystems. The above equation is only true when

$$\omega_j^{\,*}(1) = \omega_j^{\,*}(2) = ...\omega_j^{\,*}(R) = \omega_j^{\,*} \quad (12)$$

It is likely that this condition will hold in case of aircraft design.

(3) The optimum primal variables could be obtained earlier during the solution of different subsystems and it could be checked again through dual variables obtained in geometric programming.

3. MDO Model for Aircraft

Aircraft deign is a long and costly process. Traditional aircraft conceptual design establishes basic aircraft geometry and engine thrust requirements by considering basic performance constraints such as field length, speed, and altitude. Detailed assessment of more complex and realistic mission profiles that comply with air traffic and airworthiness requirements occur after the aircraft layout has been established. This can lead to sub-optimal designs and costly revisions.

The all-in-one composite aircraft design problem for range, aerodynamics, propulsion and structure can be presented as maximization of aircraft range subject to constraints on the structure, aerodynamics, propulsion, and range subsystems. These four subsystems usually have 8 coupling variables shared by each subsystem.

The coupling variables are:

1) Thickness/chord (t/c),2) Altitude (h) 3) Mach number (M), 4) Wing aspect ratio (ar_w) 5) Wing sweep angle (ang_w), 6) Wing surface area (S_{wf}) 7) Tail surface area (S_{tf}),8) Tail aspect ratio (ar_t)

These coupling variables are denoted as

$$X_n^{'} = \{t/c, h, M, ar_w, ang_w, S_{wf}, S_{tf}, ar_t\}$$

These are the most recognizable coupling variables between the subsystems; however, it is worth examining some of the following in this category: Cruise speed, empty mass, center of gravity location, drag coefficient. The variables affecting only a particular subsystem are;

Aerodynamics:

9) Tail sweep (t_{sw}) ,10) Wing lift (w_l) ,11) Tail lift (t_l)

Aerodynamics variables are denoted as

$$X_n^{''}(1) = \{t_{sw}, w_l, t_l\}$$

Propulsion: 12) Throttle (T)

The propulsion subsystem variable is denoted as $X_n^{''} = \{T\}$

Structures:13) Thicknesses (t): about 9 in all, 14) Taper ratio (λ). The structure subsystem is obviously the most important subsystem and is extremely crucial for the success of the optimization of the system and is denoted by

$$X_n^{''}(3) = \{t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8, t_9, \lambda\}$$

There are also behavioral variables which are communicated between the subsystems. These are:

15) Lift/Drag Ratio (L/D), 16) Total lifts (L) 17) Engine weight (W_e), 18) Total weight (W_t) 19) Wing twist (θ), 20) Total drag (D) 21) Fuel weight (W_f), 22) Specific fuel consumption (SFC), 23) Engine scaling factor (F_{es})

The behavioral variables communicated between the subsystems are grouped as shown below.

$$y(1) \subseteq \{L, W_t, \theta, W_f, L/D, F_{es}\}$$

This subset of behavior variables $(y(1))$ is communicated to the aerodynamics subsystem optimization problem. Whereas the following subset of the behavior variable $(y(2))$ is communicated to the structure subsystem optimization problem

$$y(2) \subseteq \{L, W_e, W_t, W_f, \theta\}$$

In total there are about 30 design variables and the constraints are nearly 45 in the total composite design. It is advisable to decompose the system in 4 subsystems as indicated earlier.

It has been found in design optimization problems that Karush-Kuhn-Tucker (KKT) conditions are not satisfied due to the high degree of non-linearity inherently involved in design problems. It may happen due to discontinuity in any of the subsystems. To

take care of such problems, a mathematical method of composing the results has been presented in the earlier section of this proposal.

The decomposition of the all-in-one problem may be accomplished in various alternate ways but the decomposition based on the coupling variables concept appears to be more suitable for the multidisciplinary design optimization of an aircraft design problem.

Multidisciplinary decomposition with a central coordinator that includes objective and R subsystems of the three disciplines will be attempted. The minimization of the cost objective will better serve the aircraft design than the range of aircraft.

Other decomposition may be attempted where the range equation acts as a coupling objective, which will eliminate the range subsystem as one of the subsystems from the optimization problem.

It is possible to solve all subsystem in parallel [16-18]. However, it is proposed to sequentially solve due to the coupling variables and some of the behavior variables. Some of the behavior variables will be required as input to some of the subsystems. This approach will need range calculations in all the subsystems. The Breguet range equation [13] may be used as a coupling objective, which will eliminate one subsystem. This approach is likely to save computation time and effort.

These approaches can be graphically shown as follows.

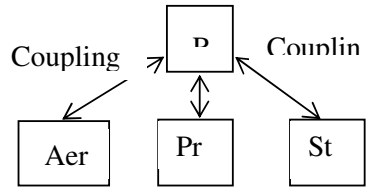


Figure 1: Traditional Approach

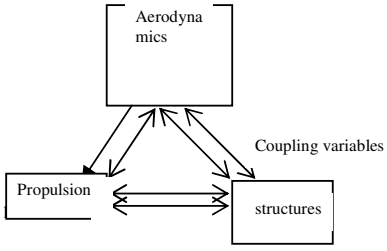


Figure 2: Alternate Approach

The formulation proposed here is conceptual rather than based on the real data from an aircraft system. Based on the design variables and coupling variables identified the subsystem optimization model can be written as below.

3.1 Coupling Subsystem:

Maximize

$$C_0^{\,0}(X_n^{\,'}) = \sum\left(\sum_{1}^{T_m} C_{0t}^{\,0}(t/c)^{a_{0t1}}(h)^{a_{0t2}}(M)^{a_{0t3}}(ar_w)^{a_{0t4}}(ang_w)^{a_{0t5}}(S_{wf})^{a_{0t6}}(S_{tf})^{a_{0t7}}(ar_t)^{a_{0t8}}\right) \quad (13)$$

Under the constraints

Upper and lower limit constraints on all the coupling variables need to be applied.

$$(t/c)_{lower} \le (t/c) \le (t/c)_{upper} \quad (14) \qquad (h)_{lower} \le h \le (h)_{upper} \quad (15),$$

$$(M)_{lower} \le M \le (M)_{upper} \quad (16),$$

$$(ar_w)_{lower} \le ar_w \le (ar_w)_{upper} \quad (17)$$

$$(ang_w)_{lower} \le ang_w \le (ang_w)_{upper} \quad (18)$$

$$(S_{wf})_{lower} \le S_{wf} \le (S_{wf})_{upper} \quad (19), \qquad (S_{tf})_{lower} \le S_{tf} \le (S_{tf})_{upper} \quad (20)$$

$$(ar_t)_{lower} \le ar_t \le (ar_t)_{upper} \quad (21)$$

Along with constraints on behavior constraints on lift/drag (L/D) ratio, total lift (L), engine weight (W_e), and wing twist (θ). There are several empirical and analytical relations available in the literature connecting these behavior variables.

$$C_1^{\,0}(X_n^{\,'}) = \sum\left(\sum_{1}^{T_m} C_{0t}^{\,0}(L/D)^{a_{0t9}}(L)^{a_{0t10}}(W_e)^{a_{0t11}}(\theta)^{a_{0t12}}\right) \quad (22)$$

$$C_2^{\,0}(X_n^{\,'}) = \sum\left(\sum_{1}^{T_m} C_{0t}^{\,0}(L/D)^{a_{0t9}}(L)^{a_{0t10}}(W_e)^{a_{0t11}}(\theta)^{a_{0t12}}\right) \quad (23)$$

$$C_3^{\,0}(X_n^{\,'}) = \sum\left(\sum_{1}^{T_m} C_{0t}^{\,0}(L/D)^{a_{0t9}}(L)^{a_{0t10}}(W_e)^{a_{0t11}}(\theta)^{a_{0t12}}\right) \quad (24)$$

$$C_4^{\,0}(X_n^{\,'}) = \sum\left(\sum_{1}^{T_m} C_{0t}^{\,0}(L/D)^{a_{0t9}}(L)^{a_{0t10}}(W_e)^{a_{0t11}}(\theta)^{a_{0t12}}\right) \quad (25)$$

.....

$$C_m^{\,0}(X_n^{\,'}) = \sum\left(\sum_{1}^{T_m} C_{0t}^{\,0}(L/D)^{a_{0t9}}(L)^{a_{0t10}}(W_e)^{a_{0t11}}(\theta)^{a_{0t12}}\right) \quad (26)$$

The exponents a_{0tm} are referring to the exponents of variables from 1to 11 involved in the coupling subsystem. Some coefficients $(C_{0t}^{\,0})$ and exponents (a_{0tm}) may be zero after developing the equations and they are either dropped or equal to 1.

3.2 Aerodynamics Subsystem:

The mathematical model of the aerodynamics subsystem [13] is presented below. The behavior and coupling variables obtained in the coupling subsystem is communicated to the aerodynamics subsystem before optimization of the subsystem is attempted. The aerodynamics discipline solves for induced drag, parasite drag, lift and drag coefficients, aerodynamic center, and stability derivatives as a function of aircraft geometry.

Maximize

$$C_1^{\,1}(X_n^{\,'}) = \sum\left(\sum_{1}^{T_m} C_{1t}^{\,1}(t/c)^{a_{1t1}}(h)^{a_{1t2}}(M)^{a_{1t3}}(ar_w)^{a_{1t4}}(ang_w)^{a_{1t5}}(S_{wf})^{a_{1t6}}(S_{tf})^{a_{1t7}}(ar_t)^{a_{1t8}}(t_{sw})^{a_{1t12}}(w_f)^{a_{1t13}}(t_f)^{a_{1t14}}\right) \quad (27)$$

Under the constraints:

The behavior constraints and lower and upper constraints on the aerodynamics variables need to be applied on this subsystem.

$$C_1^{\,1}(X_n^{\,'}(1)) = \sum\left(\sum_{i}^{opn} C_{1it}^{\,1}(t/c)^{a_{0in1}}(h)^{a_{1in2}}(M)^{a_{1in3}}(ar_w)^{a_{1in4}}(ang_w)^{a_{1in5}}(S_{wf})^{a_{1in6}}(S_{tf})^{a_{1in7}}(ar_t)^{a_{1in8}}\right)$$
$$(t_{sw})^{a_{1i12}}(w_l)^{a_{1i13}}(t_l)^{a_{1i14}}$$

(28)

$$C_2^{\,1}(X_n^{\,'}(1)) = \sum\left(\sum_{i}^{opn} C_{1it}^{\,1}(t/c)^{a_{0in1}}(h)^{a_{1in2}}(M)^{a_{1in3}}(ar_w)^{a_{1in4}}(ang_w)^{a_{1in5}}(S_{wf})^{a_{1in6}}(S_{tf})^{a_{1in7}}(ar_t)^{a_{1in8}}\right)$$
$$(t_{sw})^{a_{1i12}}(w_l)^{a_{1i13}}(t_l)^{a_{1i14}}$$

(29)

……

$$C_k^{\,1}(X_n^{\,'}(1)^{'}) = \sum\left(\sum_{i}^{opn} C_{1it}^{\,1}(t/c)^{a_{0in1}}(h)^{a_{1in2}}(M)^{a_{1in3}}(ar_w)^{a_{1in4}}(ang_w)^{a_{1in5}}(S_{wf})^{a_{1in6}}(S_{tf})^{a_{1in7}}(ar_t)^{a_{1in8}}\right)$$
$$(t_{sw})^{a_{1i12}}(w_l)^{a_{1i13}}(t_l)^{a_{1i14}}$$

(30)

The lower and upper constraints on the aerodynamics design variables should also be imposed.

$$(t_{sw})_{lower} \le t_{sw} \le (t_{sw})_{upper} \quad (31) \qquad (W_l)_{lower} \le W_l \le (W_l)_{upper} \quad (32)$$

$$(t_l)_{lower} \le t_l \le (t_l)_{upper} \quad (33)$$

Some derived constraints on static and dynamic aero elastic, thermal constraint, and control system need to be included in this formulation.

3.3 Propulsion Subsystem:

The propulsion subsystem [13] model is affected by the decisions in the coupling subsystem and aerodynamics subsystem. The coupling variables altitude (h) and Mach number (M) along with variables such as engine scaling factor (ESF), total drag (D), specific fuel consumption (SFC), and total engine weight (W_t) are communicated to the propulsion subsystem. This is the third subsystem in the MDO model.

Maximize

$$C_{10}^{\,3}{}_3(X_n^{\,'}) = \sum\left(\sum_{i}^{000} C_{0t}(h)^{a_{0t1}{}^{a_{3in}}}(M)^{a_{\omega 2}{}^{a_{0in}}}(ESF)^{a_{0i3}}(D)^{a_{0i4}}(SFC)^{a_{0i5}}(W_t)^{a_{0i6}}(T)^{a_{0i7}}\right) \quad (34)$$

Under the constraints:

The lower and upper bound constraint on the all design variables should be applied on the propulsion subsystem optimization. Then the other behavior constraints on the strength, dynamic requirements, stability and control should be imposed on the propulsion subsystem in MDO.

$$C_{1_{10}}^{\,3}{}_3(X_2^{\,'}) = \sum\left(\sum_{i}^{000} C_{1t}^{\,3}(h)^{a_{1i1}{}^{a_{3in}}}(M)^{a_{1i2}{}^{a_{0in}}}(ESF)^{a_{1i2}}(D)^{a_{1i4}}(SFC)^{a_{1i5}}(W_t)^{a1t6}(T)^{a_{1i7}}\right) \quad (35)$$

$$C_2^{\,3}(X_n^{\,'}(1)) = \sum\left(\sum_{i}^{opn} C_{1it}^{\,1}(t/c)^{a_{0in1}}(h)^{a_{1in2}}(M)^{a_{1in3}}(ar_w)^{a_{1in4}}(ang_w)^{a_{1in5}}(S_{wf})^{a_{1in6}}(S_{tf})^{a_{1in7}}(ar_t)^{a_{1in8}}\right)$$
$$(t_{sw})^{a_{1i12}}(w_l)^{a_{1i13}}(t_l)^{a_{1i14}} \quad (36)$$

$$C_3^{\,3}(X_n^{\,'}(1)) = \sum\left(\sum_{i}^{opn} C_{1it}^{\,1}(t/c)^{a_{0in1}}(h)^{a_{1in2}}(M)^{a_{1in3}}(ar_w)^{a_{1in4}}(ang_w)^{a_{1in5}}(S_{wf})^{a_{1in6}}(S_{tf})^{a_{1in7}}(ar_t)^{a_{1in8}}\right)$$
$$(t_{sw})^{a_{1i12}}(w_l)^{a_{1i13}}(t_l)^{a_{1i14}} \quad (37)$$

$$(T)_{lower} \le T \le (T)_{upper} \quad (38)$$

3.4 Structural Subsystem:

The fourth component is structural subsystem [13, 14], which is obviously very critical. The values of coupling design variables such as thickness/chord (t/c), altitude (h), Mach number (M), wing aspect ratio (ar_w), wing sweep ratio, (S_{wf}), tail surf area (S_{tf}), (ar_t) from the coupling subsystem solution is communicated to the structural subsystem. The behavior variables total lift (L), engine weight (W_e), total weight (W_t), fuel weight (W_f) along with wing twist (θ) are important parameters for the structural subsystem optimization. There are about 9 thickness design variables including taper ratio (λ) which need to be determined during the structural subsystem optimization. The structural weight and center of gravity location of each component is estimated, giving aircraft weight, center of gravity, and moments of inertia. Static stability constraints are satisfied by enforcing a margin of safety between the center of gravity and aerodynamic center.

The structural subsystem is presented below.

Maximize:

$$C_0^{\,4}(X(4)) = \sum C_{0t}{}^{3}.(t/c)^{a_{or1}}(h)^{a_{0i2}}(M)^{a_{0i3}}(ar_w)^{a_{0i5}}(\Lambda_w)^{a_{0i6}}(S_{wf})^{a_{0i7}}(S_{tf})^{a_{0i8}}(ar_t)^{a_{0i8}}$$
$$(L)^{a_{0i9}}(W_e)^{a_{0i9}}(W_t)^{a_{0i10}}(W_f)^{a_{0i11}}(\theta)^{a_{0i12}}t_1^{\,a_{0i12}}t_2^{\,a_{0i13}}t_3^{\,a_{0i14}}t_4^{\,a_{0i15}}t_5^{\,a_{0i15}}t_6^{\,a_{0i16}}t_7^{\,a_{0i17}}t_8^{\,a_{0i17}}t_9^{\,a_{0i18}}\lambda^{a_{0i19}}$$

(39)

Under the constraints:

The structural subsystem is very critical for the overall optimization of the aircraft system and hence all strictest possible constraints must be imposed. The upper and lower bound constraints on all parameters along with other variables are:

$$(t_1)_{lower} \le t_1 \le (t_1)_{upper} \quad (40), \qquad (t_2)_{lower} \le t_2 \le (t_2)_{upper} \quad (41)$$

$$(t_3)_{lower} \le t_3 \le (t_3)_{upper} \quad (42), \qquad (t_4)_{lower} \le t_4 \le (t_4)_{upper} \quad (43)$$

$$(t_5)_{lower} \le t_5 \le (t_5)_{upper} \quad (44) \qquad (t_6)_{lower} \le t_6 \le (t_6)_{upper} \quad (45)$$

$$(t_7)_{lower} \le t_7 \le (t_7)_{upper} \quad (46) \qquad (t_8)_{lower} \le t_8 \le (t_8)_{upper} \quad (47)$$

$$(t_9)_{lower} \le t_9 \le (t_9)_{upper} \quad (48) \qquad (\lambda)_{lower} \le \lambda \le (\lambda)_{upper} \quad (49)$$

In addition, these behavior constraints below are applied.

$$C_1^{\,4}(X(4)) = \sum C_{1t}{}^{4}.(t/c)^{a_{1or1}}(ar_w)^{a_{1i2}}(\Lambda_w)^{a_{1i4}}(S_{wf})^{a_{1i5}}(S_{tf})^{a_{1i65}}(ar_t)^{a_{1i68}}$$
$$t_1^{\,a_{1i72}}t_2^{\,a_{1i83}}t_3^{\,a_{1i94}}t_4^{\,a_{1i10}}t_5^{\,a_{1i11}}t_6^{\,a_{1i126}}t_7^{\,a_{1i13}}t_8^{\,a_{1i14}}t_9^{\,a_{1i15}}\lambda^{a_{1i16}} \quad (50)$$

$$C_2^{\,4}(X(4)) = \sum C_{2t}{}^{4}.(t/c)^{a_{2i1}}(ar_w)^{a_{2i25}}(\Lambda_w)^{a_{2i3}}(S_{wf})^{a_{2i4}}(S_{tf})^{a_{2i5}}(ar_t)^{a_{2i6}}$$
$$(L)^{a_{2i8}}(W_e)^{a_{2i9}}(W_t)^{a_{2i10}}(W_f)^{a_{2i11}}(\theta)^{a_{2i12}}t_1^{\,a_{2i13}}t_2^{\,a_{2i14}}t_3^{\,a_{02i15}}t_4^{\,a_{02i16}}t_5^{\,a_{21i75}}t_6^{\,a_{2i18}}t_7^{\,a_{2i18}}t_8^{\,a_{2i19}}t_9^{\,a_{2i20}}\lambda^{a_{2i219}} \quad (51)$$

$$C_3^{\,4}(X(4)) = \sum C_{3t}{}^{4}.(t/c)^{a_{3i1}}(ar_w)^{a_{3i25}}(\Lambda_w)^{a_{3i3}}(S_{wf})^{a_{3i4}}(S_{tf})^{a_{3i5}}(ar_t)^{a_{3i6}}$$
$$(L)^{a_{3i79}}(W_e)^{a_{3i8}}(W_t)^{a_{3i9}}(W_f)^{a_{3i101}}(\theta)^{a_{3i11}}t_1^{\,a_{3i122}}t_2^{\,a_{3i13}}t_3^{\,a_{3i14}}t_4^{\,a_{3i15}}t_5^{\,a_{3i16}}t_6^{\,a_{3i17}}t_7^{\,a_{3i18}}t_8^{\,a_{3i19}}t_9^{\,a_{3i20}}\lambda^{a_{3i21}} \quad (52)$$

……

$$C_m^{\,4}(X(4)) = \sum C_{mt}{}^{4}.(t/c)^{a_{mi1}}(ar_w)^{a_{mi2}}(\Lambda_w)^{a_{mi3}}(S_{wf})^{a_{mi4}}(S_{tf})^{a_{mi5}}(ar_t)^{a_{mi6}}$$
$$(L)^{a_{mi7}}(W_e)^{a_{mi8}}(W_t)^{a_{0imi8}}(W_f)^{a_{mi8}}(\theta)^{a_{mi11}}t_1^{\,a_{mi12}}t_2^{\,a_{mi13}}t_3^{\,a_{mi14}}t_4^{\,a_{mi15}}t_5^{\,a_{mi18}}t_6^{\,a_{mi19}}t_7^{\,a_{0mi20}}t_8^{\,a_{mi21}}t_9^{\,a_{mi22}}\lambda^{a_{mi23}} \quad (53)$$

The above constraints are for strength, stiffness, static aerodynamic requirements, thermal requirement, and for stability and control system. There are 10+m constraints on the structural subsystem MDO module. As noted earlier some of the coefficients (C_{mt}) may be zero and that particular constraint may drop out of the system. Similarly some of the exponents (a_{mtn}) may be zero and that variable may also drop out.

4. REDUCTION AND/OR ELIMINATION OF REDUCTION AND INFEASIBLE CONSTRAINTS

In the system optimization problem all possible real world constraints need to be included. However, that makes the optimization problem complex. It should analyzed whether all these constraints developed are active and feasible, as there is a tendency to put all possible constraints in the MDO formulation, in an attempt to make the system optimization problem more realistic. This leads to the constraints sometimes conflicting, redundant or infeasible. Hence in formulation of the MDO problem, it is beneficial to examine whether all the constraints are active and feasible. An algorithm is proposed to eliminate the redundant and infeasible constraints by an approach shown below.

Considering the nature of the MDO for aircraft design, one may reasonably assume that at least one variable is present in each constraints. In this MDO formulation the lower and upper bounds on each design variables are reasonably known.

$$X(J)_{lower} \le X(J) \le X(J)_{upper} \quad (55)$$

Where J=0,1,2,3,4 for the four subsystem including coupling subsystem.

The feasible ranges of the design variables will help in analyzing the redundancy and infeasibility of the constraints in the MDO presented above.

4.1 Algorithm:

1) Plug in the upper bound values of the design variables in the constraint and determine the numerical value,

$$C_m^{\,1}(X(m))_{upper}$$

2) calculate the numerical value of the constraint by plugging in the lower bound values, $C_m^{\,1}(X(m))_{lower}$.

3) The feasible value of the constraint is known from the real world practice, $C_m^{\,1}(X(m))_{feasible}$.

4) Compare the lower bound and upper bound values of the constraint with the feasible value and decide if the constraint is redundant and/or feasible.

If $C_m^{\,1}(X(m))_{lower} \le C_m^{\,1}(X(m))_{feasible}$ (56)

then the constraint is active and feasible. If not, it is redundant and should be removed from the subsystem optimization.

5. Automatic Mathematical Modeling Concept

It is proposed to build the MDO equations automatically from a scheme presented below. To improve the design process, a multidisciplinary team [6, 7, 8, and 9] must first list all objectives and all constraints that the aircraft design needs to meet. The standard input format for coefficients for different subsystems $C_{mt}^{\,l}$, m =0, 1, 2, M, and t =1, 2, 3…..n, $\ell = 0,1,2,3$ subsystems are provided in the matrix form. It is possible that some of the $C_{mt}^{\,l}$ values are in equation form and these values are calculated and transferred to the relevant subsystem equations. After supplying these values software should be able to generate the objective function and constraints of the particular subsystem.

5.1 Development of Coefficient Data Matrix:

Table1: Coefficient Data Matrix for coupling Subsystem

t	1	2	3	4	5	n	#of terms
$C_{0t}^{\,0}$	C_{01}	C_{02}	C_{03}	C_{04}	C_{05}	C_{0n}	Coup. obj
$C_{1t}^{\,0}$	$C_{11}^{\,0}$	$C_{12}^{\,0}$	$C_{13}^{\,0}$	$C_{14}^{\,0}$	$C_{15}^{\,0}$	$C_{1n}^{\,0}$	1st cons
$C_{2t}^{\,0}$	$C_{21}^{\,0}$	$C_{22}^{\,0}$	$C_{23}^{\,0}$	$C_{24}^{\,0}$	$C_{25}^{\,0}$	$C_{2n}^{\,0}$	2nd cons
$C_{3t}^{\,0}$	$C_{31}^{\,0}$	$C_{32}^{\,0}$	$C_{33}^{\,0}$	$C_{34}^{\,0}$	C_{35}	$C_{3n}^{\,0}$	3rd cons
$C_{4t}^{\,0}$	$C_{41}^{\,0}$	$C_{42}^{\,0}$	$C_{43}^{\,0}$	$C_{44}^{\,0}$	$C_{45}^{\,0}$	$C_{4n}^{\,0}$	4th.cons
$C_{mt}^{\,0}$	$C_{m1}^{\,0}$	$C_{m2}^{\,0}$	$C_{m3}^{\,0}$	$C_{m4}^{\,0}$	$C_{m5}^{\,0}$	$C_{mn}^{\,0}$	Mth.cons

Table 2: Coefficient Data Matrix for Aerodynamics Subsystem

t	1	2	3	4	5	n	#of terms
$C_{0t}^{\,1}$	C_{01}	C_{02}	C_{03}	$C_{04}^{\,1}$	$C_{05}^{\,1}$	$C_{0n}^{\,1}$	Coup.obj
$C_{1t}^{\,1}$	$C_{11}^{\,1}$	$C_{12}^{\,1}$	$C_{13}^{\,1}$	$C_{14}^{\,1}$	$C_{15}^{\,1}$	$C_{1n}^{\,1}$	1st cons
$C_{2t}^{\,1}$	$C_{21}^{\,1}$	$C_{22}^{\,1}$	$C_{23}^{\,1}$	$C_{24}^{\,1}$	$C_{25}^{\,1}$	$C_{2n}^{\,1}$	2nd cons
$C_{3t}^{\,1}$	$C_{31}^{\,1}$	$C_{32}^{\,1}$	$C_{33}^{\,1}$	$C_{34}^{\,1}$	$C_{35}^{\,1}$	$C_{3n}^{\,1}$	3rd cons
$C_{4t}^{\,1}$	$C_{41}^{\,1}$	$C_{42}^{\,1}$	$C_{43}^{\,1}$	$C_{44}^{\,1}$	$C_{45}^{\,1}$	$C_{4n}^{\,1}$	4th.cons
$C_{mt}^{\,1}$	$C_{m1}^{\,1}$	$C_{m1}^{\,1}$	$C_{m3}^{\,1}$	$C_{m4}^{\,1}$	$C_{m5}^{\,1}$	$C_{mn}^{\,1}$	Mth.cons

Table .3 Coefficient Data Matrix for Propulsion Subsystem

t	1	2	3	4	5	n	#of terms
$C_{0t}^{\,2}$	C_{01}	C_{02}	C_{03}	$C_{04}^{\,2}$	$C_{05}^{\,2}$	$C_{0n}^{\,2}$	Coup. obj
$C_{1t}^{\,2}$	$C_{11}^{\,2}$	$C_{12}^{\,2}$	$C_{13}^{\,2}$	$C_{14}^{\,2}$	$C_{15}^{\,2}$	$C_{1n}^{\,2}$	1st cons

C_{2t}^{2}	C_{21}^{2}	C_{22}^{2}	C_{23}^{2}	C_{24}^{2}	C_{25}^{2}	C_{2n}^{2}	2nd cons
C_{3t}^{2}	C_{31}^{2}	C_{32}^{2}	C_{33}^{2}	C_{34}^{2}	C_{35}^{2}	C_{3n}^{2}	3rd cons
C_{4t}^{2}	C_{41}^{2}	C_{42}^{2}	C_{43}^{2}	C_{44}^{2}	C_{45}^{2}	C_{4n}^{2}	4th.co ns
C_{mt}^{2}	C_{m1}^{2}	C_{m2}^{2}	C_{m3}^{2}	C_{m4}^{2}	C_{m5}^{2}	C_{mn}^{2}	Mth.c ons

Table 4: Coefficient Data Matrix for Structural Subsystem

t	1	2	3	4	5	n	#of terms
C_{0t}^{3}	C_{01}	C_{02}	C_{03}	C_{04}^{3}	C_{05}^{3}	C_{0n}^{3}	Coup. obj
C_{1t}^{3}	C_{11}^{3}	C_{12}^{3}	C_{13}^{3}	C_{14}^{3}	C_{15}^{3}	C_{1n}^{3}	1st cons
C_{2t}^{3}	C_{21}^{3}	C_{22}^{3}	C_{23}^{3}	C_{24}^{3}	C_{25}^{3}	C_{2n}^{3}	2nd cons
C_{3t}^{3}	C_{31}^{3}	C_{32}^{3}	C_{33}^{3}	C_{34}^{3}	C_{35}^{3}	C_{3n}^{3}	3rd cons
C_{4t}^{3}	C_{41}^{3}	C_{42}^{3}	C_{43}^{3}	C_{44}^{3}	C_{45}^{3}	C_{4n}^{3}	4th.co ns
C_{mt}^{3}	C_{m1}^{3}	C_{m2}^{3}	C_{m3}^{3}	C_{m4}^{3}	C_{m5}^{3}	C_{mn}^{3}	Mth.c ons

5.2 Development of the Exponent Matrix:

The exponent matrix [7-9] is also input into the equations in matrix form as shown below. All the exponents will differ from subsystem to subsystem and they need to be estimated.

Table 5: Exponent Data Matrix of Coupling Subsystem

variables	1	2	3	4	5	6	t
L/D	a_{0112}	a_{029}	a_{039}	a_{049}	a_{059}	a_{069}	a_{0t9}
L	a_{0113}	a_{0210}	a_{0310}	a_{0410}	a_{0510}	a_{0610}	a_{0t10}
W_e	a_{0114}	a_{0211}	a_{0311}	a_{0411}	a_{0511}	a_{0611}	a_{0t11}
θ	a_{0115}	a_{0212}	a_{0312}	a_{0412}	a_{0512}	a_{0612}	a_{0t12}

Table 6: Exponent Data Matrix of Aerodynamics Subsystem

variables	1	2	3	4	5	6	t
t/c	a_{011}	a_{021}	a_{031}	a_{041}	a_{051}	a_{061}	a_{0t1}
h	a_{012}	a_{022}	a_{032}	a_{042}	a_{052}	a_{062}	a_{0t2}
M	a_{013}	a_{023}	a_{033}	a_{043}	a_{053}	a_{063}	a_{0t3}
ar_w	a_{014}	a_{024}	a_{034}	a_{044}	a_{054}	a_{064}	a_{0t4}
ang_w	a_{015}	a_{025}	a_{035}	a_{045}	a_{055}	a_{065}	a_{0t5}
S_{wf}	a_{016}	a_{026}	a_{036}	a_{046}	a_{056}	a_{066}	a_{0t6}
S_{tf}	a_{017}	a_{027}	a_{037}	a_{047}	a_{057}	a_{067}	a_{0t7}
ar_t	a_{018}	a_{028}	a_{038}	a_{048}	a_{058}	a_{068}	a_{0t8}
t_{sw}	a_{019}	a_{029}	a_{039}	a_{049}	a_{059}	a_{069}	a_{0t9}
W_l	a_{011}	a_{0210}	a_{0310}	a_{041}	a_{051}	a_{0610}	a_{0t10}
t_l	a_{011}	a_{0211}	a_{0311}	a_{041}	a_{051}	a_{0611}	a_{0t11}

Table 7: Exponent Data Matrix of Propulsion Subsystem

variables	1	2	3	4	5	6	t
T	a_{0116}	a_{0216}	a_{0316}	a_{0416}	a_{0516}	a_{0616}	a_{0t16}

Table 8: Exponent Data Matrix of Structural Subsystem

variables	1	2	3	4	5	6	t
t_1	a_{011}	a_{0217}	a_{0317}	a_{0417}	a_{0517}	a_{0617}	a_{0t17}
t_2	a_{011}	a_{0218}	a_{0318}	a_{0418}	a_{0518}	a_{0618}	a_{0t18}
t_3	a_{011}	a_{0219}	a_{0319}	a_{0419}	a_{0519}	a_{0619}	a_{0t19}
t_3	a_{012}	a_{0220}	a_{0320}	a_{0420}	a_{0520}	a_{0620}	a_{0t20}
t_4	a_{012}	a_{0221}	a_{0321}	a_{0421}	a_{0521}	a_{0621}	a_{0t21}
t_5	a_{012}	a_{0222}	a_{0322}	a_{0422}	a_{0522}	a_{0622}	a_{0t22}
t_6	a_{012}	a_{0223}	a_{0323}	a_{0423}	a_{0523}	a_{0623}	a_{0t23}
t_7	a_{012}	a_{0224}	a_{0324}	a_{0424}	a_{0524}	a_{0624}	a_{0t24}

t_8	a_{011}	a_{0225}	a_{0325}	a_{0425}	a_{0525}	a_{0625}	a_{0r25}
t_9	a_{011}	a_{0226}	a_{0326}	a_{0426}	a_{0526}	a_{0626}	a_{0r26}
λ	a_{011}	a_{0227}	a_{0327}	a_{0427}	a_{0527}	a_{0627}	a_{0r27}

7. SOLUTION

The coefficients and exponents shown in the above matrices are theoretical as stated above. The real values of these are transferred to a different subsystem solution of the coupling subsystem and then it will proceed to aerodynamics, propulsion, and finally to the structural subsystem sequentially.

As stated earlier, an augmented geometric program [3-6] is followed to reformulate the subsystem mathematical model before solution. In certain cases, the degree of difficulty is reduced in this formulation. It has been found that this approach also helps to overcome the problem of "loose" constraints. It is likely to happen when the spread of lower and upper range of variables are large. This principal idea is to create a new augmented program equivalent to the original subsystem program by adding slack variables to all constraints, then executing or solving this equivalence by geometric programming methodology. It has been found that if a solution vector X'_n exists, the constraints on structural requirements, static and dynamic aeroelastic requirements, thermal environment, and other relevant constraints will all be tight at the optimum. The problem of degeneracy and consistency can also tackled at this stage.

The TOMLAB software, from Tomlab Optimization Inc. [15] will be used for the solution. Tom lab can solve geometric programs for large degrees of difficulty. It will be of interest to watch the values of the optimized design variables in different subsystems. If there is any discrepancy in some optimized design values it is likely to be due to non-smoothness of some constraints in different subsystems. The main objective of the proposal is to make an attempt to present a robust mathematical programming method for the formulation and solution of multidisciplinary design optimization (MDO) problems. The method presents a smart decomposition technique with coupling subsystem and coupling variables. The large degree of difficulty which is a curse of geometric programming can be tackled also through flexible decomposition. Finally the system optimum range value can be obtained as;

$$C = \text{the maximum value of the complete system} = \left(\sum_{l=1}^{R} (C^{\ell}{}_0 (X(l)))^{1/\tau} \right)^{\tau} \quad (57)$$

Where $\tau = 1 - \sum \omega_j^*$ (58) and ω_j^* are optimum dual variables associated with coupling subsystems. The above equation is only true when

$$\varpi_j{}^*(1) = \varpi_j{}^*(2) = \ldots\ldots \varpi_j{}^*(R) = \varpi_j{}^* \quad (59)$$

It is hoped that this condition will hold in case of aircraft design. In geometric programming dual variables are obtained during the solution phase.

8. CONCLUSION

The proposal being presented looks into the problem of formulation for large multidisciplinary optimization design problems. A mathematically rigorous method for formulation and solution of MDO has been developed in this proposal. It also presents a new concept of automatic mathematical modeling, which if perfected, can be generalized and made easier to use and apply MDO to a large number of complex and complicated engineering design problems. Furthermore, a concept of reduction and elimination of redundant constraints is presented. It seems that during the development of complicated large-scale design problems the engineers try to include all types of constraints which may or may not be practical or useful. Hence, there is a need for an algorithm to eliminate some of the redundant constraints, wherever possible. The hope is that by integrating the reduction/elimination algorithm some of these problems can be eased. The methodology presented includes both the coupling and behavioral variables which move back and forth between different subsystems during the solution process. The solution is based on the augmented geometric program. The interactions and interrelation between different subsystems would be preserved in MDO by this approach. The algorithm is generic and not related to a specific problem. However, it must be stated that it is more methodological and conceptual or idea-based.

9.REFERENCES

[1] Ilan, Kroo, Valarie Manning, " Collaborative Optimization: Status and Directions," Presented at the 8th AIAA/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, 6-8 September 2000, Long Beach CA.

[2] Sobieszczanski,J.,Agte J.S.,and JrR.R.Sandusky,'Bi-Level Integrated System Synthesis (BLISS)," Technical Report NASA-AIAA-98-4916

[3] Avriel, M. and Wilde, D.J., " Stochastic Geometric Programming," Proceedings of Princeton Symposium on Mathematical Programming, H.W. Kuhn editor, Princeton University Press, 1970.

[4] Wiebking, R.D., "Deterministic and Stochastic Geometric Programming Models for Optimal Engineering Design Problems in Electric Power Generation and Computer Solutions," Ph.D. Thesis Rensselaer Polytechnic Institute, Troy, New York (1974).

[5] Jha, Nand K., "Stochastic Mathematical Modeling and Manufacturing Cost Estimation in Uncertain Industrial Environment," International Journal of Production Research, Vol.12,2755-2771, 1992.

[7] Raymer, Daniel, Aircraft Design: A Conceptual Approach, Fourth Edition, AIAA,2006

[8] Jha, Nand K., "A Discrete Data Base Multiple Objective Optimization of Milling Operation Through Geometric Programming," ASME Transaction, Journal of Engineering for Industry, Vol. 112, No.4, 1990.

[9] Jha, Nand K., "Multidisciplinary Automatic Optimal Design of Milling Cutter," Proc. Of the Third International Conference on Metal Cutting and High Speed Machining, edited by D. Dudzinski, A. Molinari, and H. Schulz, 2001, Metz, France.

[10] Jha Nand K., "Mathematical Programming Formulation and Optimal Design of Robot Control System," Applied Mathematical Modeling, 1996, Vol.20, November, pp 853-865.

[11]]Jha Nand K., "Probabilistic Cost Estimation in Advance of Production Through Stochastic Geometric Programming," Computers & Industrial Engineering, Vol.30, No4, pp. 809-821, 1996.

[12] Alyaqout SF, Papalambros PY, Ulsoy AG, "Quantification and use of system coupling in decomposed design optimization problems", Proc. ASME International Mechanical Engineering Congress and Exposition, Orlando, Florida, November 2005.

Numerical Study of Microchannel Liquid Cooling under Uniform and Non-uniform Heating Conditions

Ling Ling **Yanfeng Fan** **Ibrahim Hassan**
Department of Mechanical and Industrial Engineering, Concordia University, Montreal, Quebec, Canada H3G1M8
ling_lin@encs.concordia.ca f_yanfen@encs.concordia.ca ibrahimh@alcor.concordia.ca

Abstract

Microchannel heat sink is a promising cooling technology for semiconductor devices with a high cooling capacity and low mass flow rate requirement. Nowadays, semiconductor devices are widely used in industry and improved in size and clock speed which leads to higher heat flux. In this paper, three kinds of microchannel heat sink designs are proposed and the cooling performance is investigated by CFD. Uniform heating condition is applied to the swirl microchannel heat sink. The non-uniform heating condition is divided into local high heat flux and continuously varying heat flux to simulate the electrical devices and concentrated photovoltaic cell, respectively. The cross-linked microchannel heat sink and hybrid micro heat sink are used to remove the local high heat flux. Two different sizes and numbers of hotspots are designed and applied to these two heat sinks. The straight microchannel heat sink is designed for continuously varying heat flux. Water is used as coolant and copper/ silicon is selected as heat sink material. The heat flux varies from 10 W/cm^2 to 100 W/cm^2. The inlet velocity varies from 0.5 m/s to 1 m/s. The maximum temperature, temperature gradient, pressure drop and total thermal resistance are chosen to be criterion of evaluating their cooling performance.

Keywords: CFD, microchannel heat sink, straight channel, cross-linked channel, swirl channel, uniform and non-uniform heating

1. Introduction

The semiconductor technology and micro fabrication method develop fast and has changed our lives dramatically. The applications of semiconductor technology have been widely extended into many areas including civil products, military and alternative energy and so on. Nowadays, the semiconductor industry is facing the problem of fast removal of high heat flux from the microchip and the system since the high operation temperature leads to the reduction of reliability, lifespan and speed [1]. Today the heat dissipation rate from electronic chips is touching 200 W/cm^2 [2]. The future cooling target is to maintain the maximum temperature of micro electrical devices lower than 85 °C and the spatial temperature variation below 10 °C [3].

Liquid cooling in microchannels was first proposed by Tuckerman and Pease [4] and it showed high heat dissipation up to 790 W/cm^2. Much work has been done due to this attracting capacity. According to the wall boundary conditions, the study of microchannel heat sinks can be categorized into uniform heating [5-8] and non-uniform heating [9-11]. Uniform heating condition is widely studied numerically and experimentally. Chen et al. [5] simulated three shapes of cross-section of microchannel: triangular, rectangular, and trapezoidal. The constant solid and fluid properties were assumed. The results showed that the triangular microchannel heat sink had best cooling performance. Lelea [6] numerically investigated a microchannel heat sink with a circular cross-section. The inlet channel had an offset of quarter of the tube perimeter from the center of micro tube in order to generate the swirl fluid. The water had temperature-dependent properties. The results showed that a strong swirl flow pattern was created and was able to enhance the heat transfer at high mass flow rate. Xu et al. [7-8] conducted both experimental and numerical study of cross-linked microchannel heat sink performance. Under uniform heating condition, results proved the existence of thermal boundary redeveloping at the exit of cross-linked channels. Heat transfer was significantly enhanced due to the presence of cross-linked channels. Pressure drop was discovered to be similar or slightly reduced for cross-linked channel heat sink.

Non-uniform heating conditions are recently studied because it is more practical in real applications. It can be classified as local high heat flux (hot spot) and continuously varying heat flux according to the applications. The typical application of hot spot is the computational devices such as CPU. The cooling performance of different microchannel heat sinks under the non-uniform heating condition has been investigated. Zhang et al. [9] simulated a microchannel heat sink under both uniform and non-uniform heating conditions. A new type of heat sink with variable channel width was developed. With narrower and more channels above the hotspot, the heat dissipation rate increased about 10%. Lee [10] numerically optimized the heat sink design used for liquid cooling IGBT power module with six dies. Different layouts of square pins directly mounted under the hot spots were designed to find out the optimization. The results showed that the maximum junction temperature was maintained at 100 °C with a spatial temperature variation of 1 °C. Chauhan et al. [11] numerically investigated the cooling performance of parallel straight channels. The processor, Intel® Pentium® 4 Northwood-style datapath with a 3 GHz clock, was selected as heat source. The chip was divided into 24 components with different heat fluxes. Five different plans were proposed based on the original floor plan. The high heat flux components were arranged close to the inlet and far from each other. The results showed that the maximum temperature of new floor plan reduced by 4.4 °C. The thermal condition of continuously varying

heat flux can be found in the application of concentrated solar photovoltaic cells caused by the concentrator or reflector. Kermani et al. [12] proposed a microchannel concentrated solar cell cooling design with manifold. The concept of this design was to redevelop the thermal boundary layer in order to obtain the high heat transfer coefficient at each channel. The experiment was achieved to investigate the pressure drop and cooling performance at the heat flux of 75 W/cm^2, which simulated a high concentrated solar cell with 1000 suns. The experimental results showed that the heat transfer coefficient reached 65 kW/K·m^2 at the flow rate of 1 g/s with a pressure drop of 15 kPa.

In summary, the microchannel heat sinks have a promising capacity of heat dissipation from a micro scale device. It has large potential to satisfy the future cooling demand. In this paper, four microchannel heat sinks have been designed and studied numerically under uniform and non-uniform heating conditions. Various heat fluxes and mass flow rates are tested. The maximum temperature, temperature variation and pressure drop at different operation conditions are presented.

2. Test Model Designs

Swirl heat sink under uniform heating condition

Swirl microchannel heat sink is designed, as shown in Fig.1a, to dissipate the large heat flux from the semiconductor devices. The present heat sink applies multiple swirl microchannels positioned in a circular flat plate to enhance the heat convection by creating the secondary flow at high Reynolds numbers. The channel depth and width are fixed as 0.5 mm and 0.4 mm, respectively. The heat is injected into the system from the bottom of heat sink at the uniform heat fluxes from 10 to 60 W/cm^2. Flow is supplied from the top of micro heat sink through a jet hole with a diameter of 2 mm and enters swirl microchannels at the volume flow rates varying from 47 to 188 ml/min.

Cross-linked heat sink with two local high heat fluxes

The non-uniform heating condition is applied to the cross-linked microchannel heat sink. Three heat sinks with different widths of cross-linked channels locating above the center of hotspots are studied and compared to conventional straight microchannel heat sink. Half of the module is chosen to be the computational domain, which is shown in Fig.1b. Two hotspots of 10 mm × 2 mm are placed at the bottom surface. First, the four heat sinks are tested under single hotspot condition, upstream hotspots (HCU) or downstream hotspots (HCD). Then, the heat sinks are subjected to both hotspots. Heating condition 1 (HC1), stands for heating condition with 100 W/cm^2 heat flux on upstream hotspots, and 50 W/cm^2 heat flux on downstream hotspots. Heating condition 2 (HC2) is the converse heating condition of HC1. Two inlet velocities, 0.5 m/s and 1 m/s, are tested for each heat sink.

Hybrid micro heat sink with six local high heat fluxes

The micro heat sink applying the jet-impingement and cross flow is designed to dissipate the heat from the electrical devices. The structure is shown in Fig.1c. Six hotspots of 2 mm × 2 mm are positioned on a flat plate of 25.4 mm × 25.4 mm. The area of flat plate except the hotspots is provided a constant heat flux of 20 W/cm^2 as background heating source among cases. Four heat fluxes from 40 to 100 W/cm^2 on the hotspots are tested to simulate the different operation conditions. The cross flow is used to remove the background heat flux and jet flow is supplied into the swirl microchannel, located at the right top of hotspot, to dissipate the large heat flux from hotspots. Due to the increased temperature of cross flow along the flowing direction, the circular pin fins are placed in the heat sink to enlarge the heat transfer area. The diameter of circular pins reduces and more pins are added along the stream wise direction to break the thermal boundary layer and obtain convective heat transfer area so that increase the heat transfer coefficient. The channel depth is 0.5 mm and the width of swirl microchannel is 0.38 mm. The cross flow and jet flow velocity vary from 0.1 m/s to 0.5 m/s and from 0.5 m/s to 2 m/s, respectively.

Straight channel heat sink with continuously varying heat flux

The continuously varying heat flux, which appears on the application of concentrated photovoltaic cells, is applied on the bottom surface of straight channel heat sinks. Two different layouts of micro heat sink on the heat source are proposed. The computational domains are ½ and ¼ of the whole module for layout 1 and layout 2, respectively. In order to simulate the non-uniform heating boundary condition, the illuminated surface in the model has been divided into multiple zones. The sketch of simulation method is shown in Fig.2. The heat flux distribution curve is divided and the average value of heat flux in each zone can be determined by,

$$\dot{q}_i = \frac{\int_{x_i}^{x_{i+1}} \dot{q}(x)dx}{W_i} \quad i = 1,2,3\ldots N \qquad (1)$$

where,

$$W_i = x_{i+1} - x_i$$

where, $\dot{q}_i$ is the average value of heat flux in the i^{th} zone; W_i is the width of the i^{th} zone; $\dot{q}(x)$ is the heat flux generated by the concentrator; x_i and x_{i+1} are the left and right coordinates of the i^{th} zone. The error between the simulated and practical heating condition can be evaluated by,

$$E_i = \left| \frac{\dot{q}_i - \dot{q}(x)}{\dot{q}_i} \right|$$

where, E_i is the error in the i^{th} zone. In this work, the error in all zones is maintained below 15%. Hence, as soon as the function of heat flux is, $\dot{q}(x)$, known, the width of each zone can be determined by solving the equation,

$$E_i = \left| \frac{\int_{x_i}^{x_{i+1}} \dot{q}(x)dx - \dot{q}(x)}{\dot{q}_x} \right| = 15\% \pm 1\% \qquad (2)$$

The value of x_1 is set as zero since the first zone starts from the center. Then, the value of x_2 is calculated based on Eq. (2). All coordinates then can be obtained using this method. In heat sink layout 1, the heat flux decreases along the x axis, while in heat sink layout 2, heat flux decreases along the flow direction, as shown in Fig.2. The heated surface is underneath the channels with an area of 10.75 mm × 10 mm. The maximum heat flux is fixed at 100 W/cm^2 in two tested heating cases. The case of q1 has a minimum heat flux of 20 W/cm^2 and a total power 18.7 W on the bottom surface of the whole module; while the minimum heat flux and total power of q2 are 50 W/cm^2 and 31.8 W, respectively. The tested inlet mass flow rate varies from 4.5 g/min to 86.1 g/min.

3. Numerical Study

The commercial software Fluent is applied to simulate this model and evaluate the cooling performance. Some assumptions are proposed before doing the simulation. Liquid has temperature-dependent thermophysical properties, such as density, viscosity, conductivity and thermal capacity; solid material has constant properties; the flow can be viewed as steady-state; no-slip boundary conditions exist at inner walls; body force is neglected due to micro-scale dimension. According to the assumptions, the governing equations that contain continuity equation, Navier-Stokes equation and energy equation are represented as:

$$\nabla(\rho \mathbf{V}) = 0 \tag{3}$$

$$\mathbf{V} \cdot \nabla(\rho \mathbf{V}) = -\nabla \mathrm{P} + \nabla(\mu \nabla \mathbf{V}) \tag{4}$$

$$\mathbf{V} \cdot \nabla(\rho C_p T_f) = \nabla(k_f \nabla T_f) \text{ (Liquid)} \tag{5}$$

$$k_w \nabla^2 T_w = 0 \text{ (Solid)} \tag{6}$$

The Navier-Stokes equation and continuity equation can be solved to get velocity in case of an incompressible fluid. The velocity is substituted into energy equation to solve the temperature distribution. Water and copper/silicon are defined as coolant and the heat sink material. The temperature-dependent properties of liquid are listed in Table.1. The operating pressure is fixed at 10^5 Pa. The "velocity-inlet" boundary condition is assigned at the inlet. The velocity direction is normal to the inlet. The temperature keeps a constant value at the inlets. The outlet boundary is set at "pressure-outlet". The gage pressure is fixed at zero at the outlet. The "SIMPLE" module is applied for pressure-velocity coupling analysis of flows in microchannels. The discretization is "second-order upwind" for velocity and energy, and "second-order upwind" for pressure. The number of iteration is set at 3000 and the convergence criterion is 10^{-6} for continuity, velocity, and 10^{-8} for energy in order to get stable results. The grid independence is performed in advance.

4. Results and Discussion

Swirl heat sink under uniform heating condition

Three different curvatures of swirl microchannels are selected and the results show that the larger curvature has better cooling performance as shown in Fig.3a since the stronger secondary flow is generated and the heat transfer is enhanced. Figure 3b shows the effect of channel number on the cooling performance. Increasing the channel number is able to decrease the temperature eve if the flow rate in each channel is reduced, which is caused by the increased contact area. The liquid removes the heat more uniformly as the channel number increases since the channel covers larger area. At the same flow rate, the pressure drop reduces as the channel number increases since the velocity in each microchannel is reduced as shown in Fig.3c. Figure 3d presents the thermal resistance of swirl microchannel heat sink. The total thermal resistance increases as the heat flux since the cooling capacity is reduced at the high heat flux. Also, increasing the velocity can enhance the cooling, which results in the smaller thermal resistance.

Non-uniform heating condition

Local high heat flux (Hot spot)

Two designs are proposed and investigated under local high heat flux conditions: cross-linked microchannel heat sink and hybrid micro heat sink. Different sizes and locations of hotspots are applied in two designs, respectively.

Cross-linked microchannel heat sink

It is found out that all heat sinks have lower maximum temperature and maximum temperature difference when hotspot is placed in the upstream region or upstream hotspot is subjected to a higher heat flux than the downstream one, as shown in Fig.4a. The temperature profile of the center lines of hotspots are presented in Fig.4b. With the presence of cross-linked channels, temperature distribution has better uniformity at the hotspot zones and pressure drop of the heat sink decreases (Fig.4c). Figure 4d compares the total thermal resistance of all cases. In conclusion, the cross-linked heat sink has the best cooling performance with a cross-linked channel width of 0.5 mm. Cross-linked channels reduced total thermal resistance when cross-linked channel width W_{cr} = 0.5 mm; while further increase in cross-linked channel width results in higher total thermal resistance and smaller cooling capacity. Thermal resistance is lower under HC1 than that under HC2, namely, the heat sink has higher cooling capacity when applying higher heat flux at HS1.

Hybrid micro heat sink

Jet impingement and cross flow are applied to dissipate the heat from the hotspots and background, respectively. The effects of cross flow and jet flow on the cooling performance are studied. The maximum temperature, temperature variation and the thermal resistance are presented here, as shown in Fig.5a-c. In the tested range, the maximum temperature is below 337.5 K and the temperature variation is lower than 10 K, which could satisfy the cooling requirement of micro electrical devices. The maximum pressure drop of 5.5 kPa occurs at the jet inlet when the jet velocity is 2 m/s. Figure 5c shows the overall thermal resistance in the tested range. The thermal resistance becomes smaller as jet flow

velocity increases since faster jet flow can remove large heat from the base and reduce the maximum temperature. Furthermore, the thermal resistance at high hotspot heat flux is smaller than one at low hotspot heat flux. Therefore, hybrid micro heat sink is suitable for micro electrical devices with high heat flux.

Continuously varying heat flux

The new method is proposed to simulate the thermal boundary of concentrated photovoltaic cell. The straight microchannel heat sink is applied to complete this task. Two orientations are tested. The maximum temperature, temperature difference, pressure drop and total thermal resistance are shown in Fig.6. In general, heat sink layout 2 has better cooling performance than layout 1. The maximum temperature, temperature difference and total thermal resistance decrease as increasing inlet mass flow rate or reducing heat flux. All three parameters have a critical value of mass flow rate. Further increasing the mass flow rate has little effect on them. The pressure drop rises with increase of mass flow rate. Furthermore, the pressure drop increases much drastically for layout 1 than that of layout 2.

5. Conclusion

In order to remove high heat flux from semiconductor devices, three kinds of heat sinks are designed and their cooling performances are numerical studied. The uniform and non-uniform heating conditions are applied. For the swirl heat sink which works under uniform heat flux, large values of channel number and curvature enhance heat dissipation rate. The temperature variation occurs from center to edge. In the cross-linked microchannel heat sink designed to cool two hotspots, the presence of cross-linked microchannels improves temperature uniformity on the bottom surface of the heat sink. The heat sink with jet-impingement shows satisfactory performance for removing six local high heat flux spots. Under continuously varying heat flux condition, the heat sink can obtain a much lower maximum temperature and temperature difference when the heat flux decreases along the flow direction. All proposed designs show satisfying cooling performance. The cooling capacity is able to remove the heat flux up to 60 W/cm^2 with a temperature variation lower than 16 K under both uniform and non-uniform heating conditions. The microchannel liquid cooling has a large potential to remove the high heat flux from the semiconductor devices in the future applications.

6. Reference

[1] Schmidt, R., and Notohardjono, B. D., 2002, "High-End Server Low-Temperature Cooling", **Ibm Journal of Research and Development**, 46, pp.739-751.

[2] Watwe, A, and Viswanath, R., 2003, "Thermal Implications of Non-Uniform Die Power Map and CPU Performance", **Proceedings of InterPACK'03**, 35044.

[3] Bar-Cohen, A., Arik, M., and Ohadi, M., 2006, "Direct Liquid Cooling of High Flux Micro and Nano Electronic Components", **Proceedings of the IEEE**, 94, pp.1549-1570.

[4] Tuckerman, D. B., and Pease, R. F. W., 1981, "High-Performance Heat Sinking for VLSI", **IEEE Electron Device Letters**, EDL-2, pp. 126-129.

[5] Chen, Y., Zhang, C., Shi, M., and Wu, J., 2009, "Three-Dimensional Numerical Simulation of Heat and Fluid Flow in Noncircular Microchannel Heat Sinks", **International Communications in Heat and Mass Transfer**, 36, pp. 917-920.

[6] Lelea, D., 2009, "The Microtube Heat Sink with Tangential Impingement Jet and Variable Fluid Properties", **Heat and Mass Transfer/Waerme-und Stoffuebertragung**, 45, pp. 1215-1222.

[7] Xu, J. L., Gan, Y. H., Zhang, D. C., and Li, X. H., 2005, "Microscale Heat Transfer Enhancement using Thermal Boundary Layer Redeveloping Concept", **International Journal of Heat and Mass Transfer**, 48, pp. 1662-74.

[8] Xu, J. L., Song, Y. X., Zhang, W., Zhang, H., and Gan, Y. H., 2008, "Numerical Simulations of Interrupted and Conventional Microchannel Heat Sinks", **International Journal of Heat and Mass Transfer**, 51, pp. 5906-5917.

[9] Zhang, M., Wang, X., Liu, H., Wang, G., 2008, "Numerical Simulation of the Micro-Channel Heat Sink on Non-Uniform Heat Source", **2008 International Conference on Electronic Packaging Technology & High Density Packaging (ICEPT-HDP).**

[10] Lee, T. T., 2000, "Design Optimization of an Integrated Liquid-Cooled IGBT Power Module using CFD Technique", **IEEE Transactions on Components and Packaging Technologies**, 23, pp. 55-60.

[11] Chauhan, A., Sammakia, B., Ghose, K., Refai-Ahmed, G., and Agonafer, D., 2010, "Hot Spot Mitigation using Single-Phase Microchannel Cooling for Microprocessors", **2010 12th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm),** pp.1-11.

[12] Kermani, E., Dessiatoun, S., Shooshtari, A., and Ohadi, M. M., "Experimental Investigation of Heat Transfer Performance of a Manifold Microchannel Heat Sink for Cooling of Concentrated Solar Cells", **2009 IEEE 59th Electronic Components and Technology Conference (ECTC 2009)**, pp. 453-9.

Table 1: Polynomial coefficient of thermal-dependent properties

	A_1	A_2	A_3	A_4	A_5
ρ	1227.8	-3.0726	0.011778	-1.56E-05	
k_f	-1.0294	0.010879	-2.26E-05	1.54E-08	
C_p	4631.9	-1.478	-3.108E-3	1.11E-05	
μ	0.33158	-3.752E-3	1.60E-05	-3.06E-08	2.19E-11

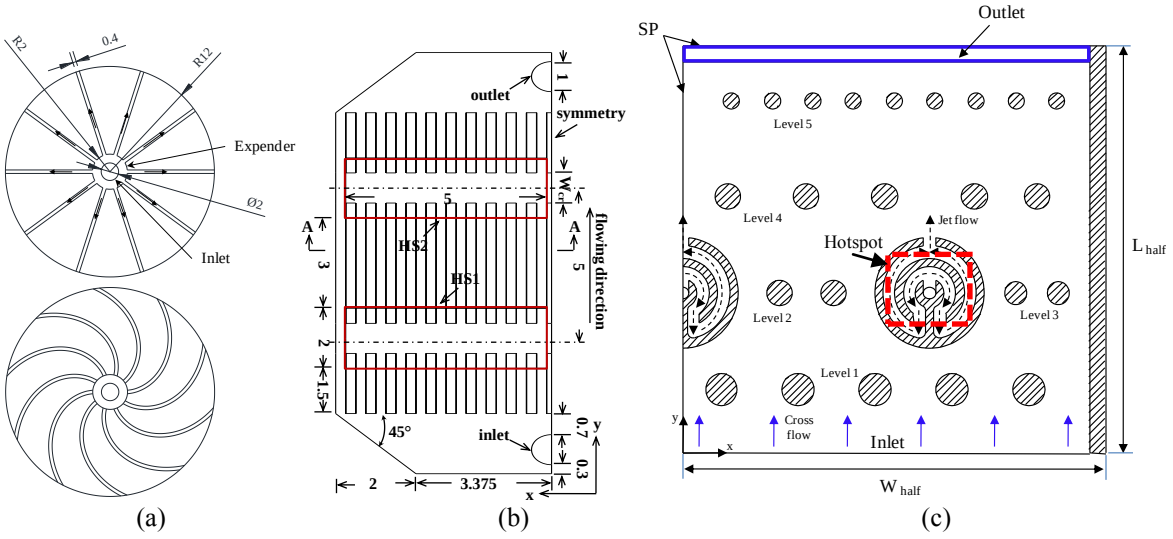


Fig.1 the structure of microchannel heat sinks (a) swirl microchannel heat sink (b) cross-linked microchannel heat sink (c) hybrid micro heat sink (unit: mm)

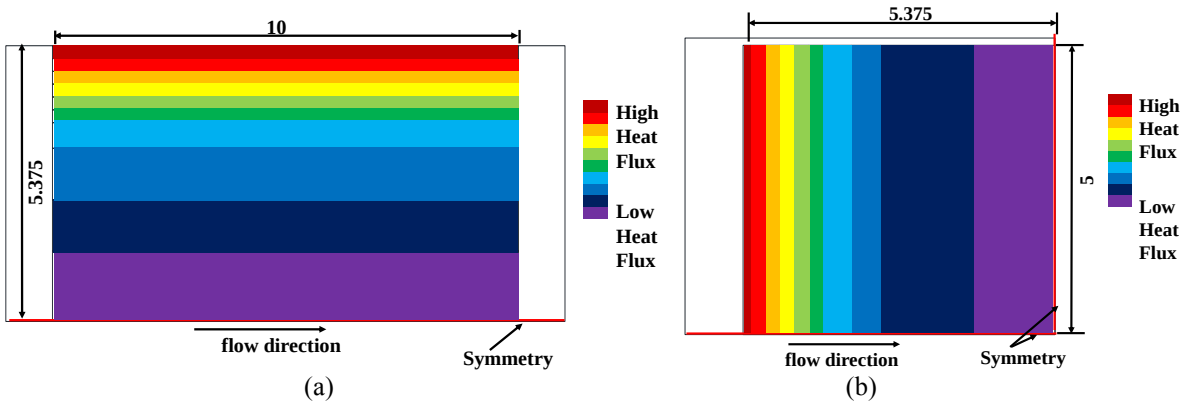


Fig.2 simulation method of heat flux (a) layout 1 (b) layout 2 (unit: mm)

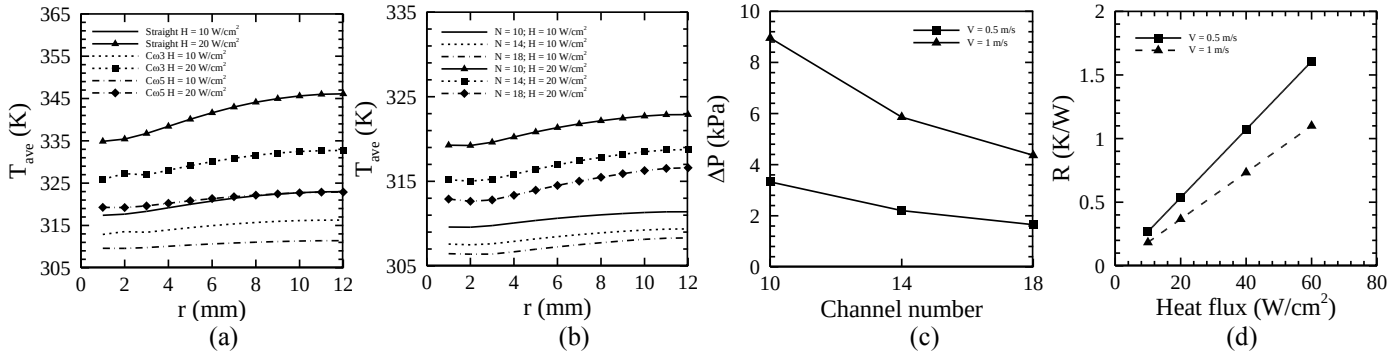


Fig.3 the numerical results of swirl microchannel heat sink (a) averaged temperature along radial direction in heat sinks with different curvatures (b) averaged temperature along the radial direction in heat sinks with different channel number (c) pressure drops of heat sinks with different channel numbers (d) thermal resistance at different heat flux and inlet velocities

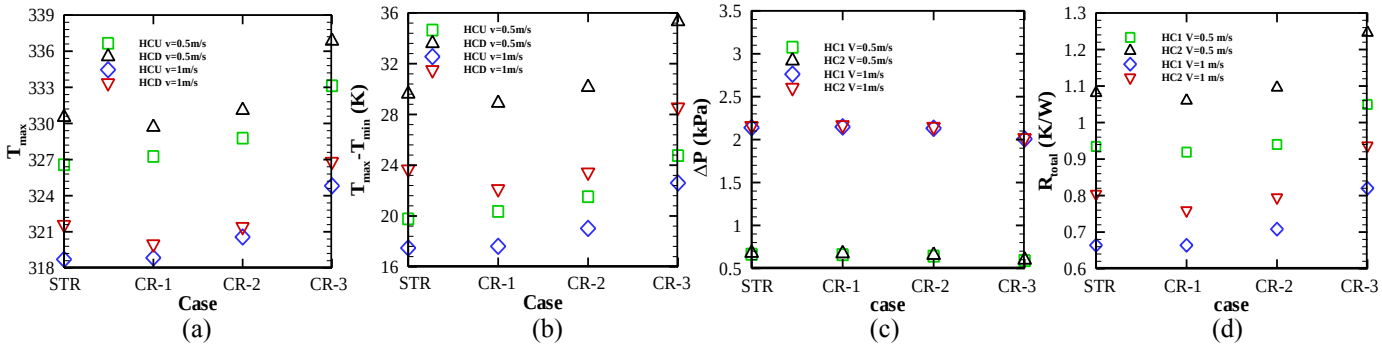


Fig.4 the results of cross-linked microchannel heat sink at different cases

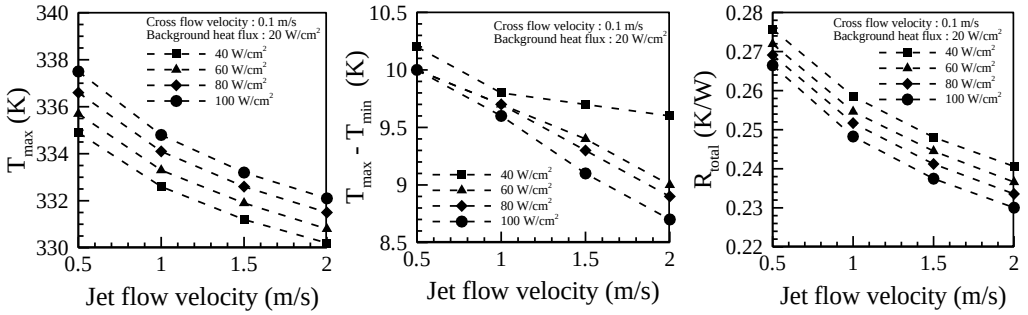


Fig.5 the results of Hybrid micro heat sink at different operation conditions

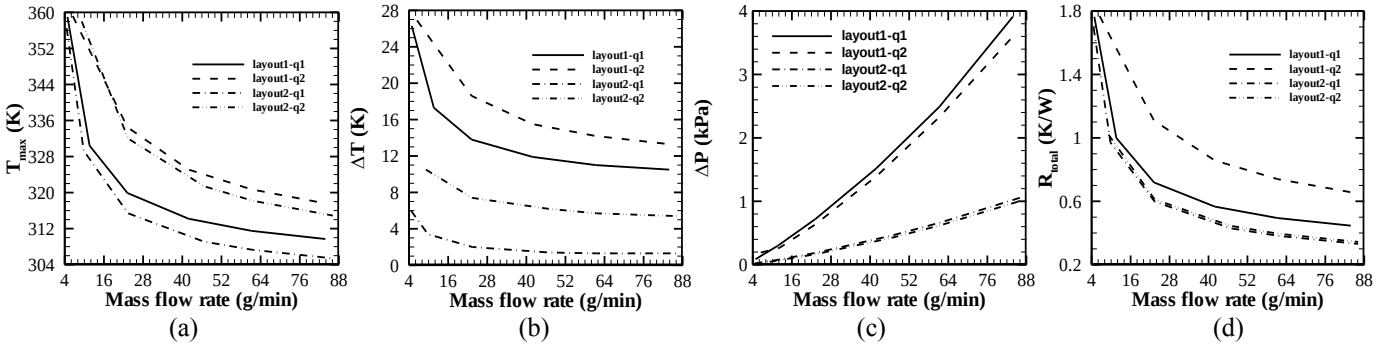


Fig.6 the results of straight microchannel heat sink at different mass flow rates

Physics of Transient Stall on a Rotating Blade Due to Blockage

Vrishank Raghav

And

Narayanan M Komerath

School of Aerospace Engineering, Georgia Institute of Technology

Atlanta, GA 30332, USA

ABSTRACT

Rotating blades encounter a blocked flow in many instances, for example the tower wakes of downwind wind turbine blades, the blades of the radiator fan of an automobile. One of the two schools of thought models the blocked flow as causing a transient stall on the blade and the other models it as a lift deficiency function. In this work the extreme case of a blocked flow is studied, where the blockage covers a considerable sector of the rotating disc. The blade undergoes a transient stall and consequently the radial flow over the blade becomes significant. Pulsed laser sheet imaging and particle image velocimetry (PIV) are used to verify the occurrence of separation, and to capture features of the separated flow. The results verify the occurrence of stall in the obstructed-inflow sector. In the stalled region, radial flow develops, but does not increase on moving outboard. This appears to be due to the presence of separation cells along the radius, which contribute to exchange of fluid between the near-blade surface and the freestream.

Keywords: Transient Stall, Rotating Blade, Blockage, and Radial Flow.

1. INTRODUCTION

Rotating blades experience varying dynamic loads due to blockage effects. Blockage is defined as a flow that is obstructed from freely interacting with the rotating blade. This dynamic loading is attributed to the blade transient stall as a result of its interaction with the blocked flow. The underlying source of the stall being the pitch increase on the blade in the low dynamic pressure region coupled with rapidly changing inflow velocity, causing the effective angle of attack to exceed the static stall limit rapidly and substantially. This triggers a sharp dynamic lift increase followed by a sudden loss of lift, and the stall persists until the blade is well under the static stall limit. As such this phenomenon has been studied extensively, two dimensional dynamic stall of 2-D airfoils has been well studied by McCroskey (1). Concepts for stall alleviation by Carr (2) and elimination by McAlister (3) have been developed for rotating blades. But much less is known about the complexities and intricacies of 3-D dynamic stall, due to the presence of strong radial acceleration.

On the other hand, Leishman (4) postulates that the dynamic lift acting on a wind turbine blade when passing through the tower wake can be modeled by fitting a lift deficiency function to the airfoil section. He suggests using a Kussner Function, which models a vertical gust impingement onto the airfoil section. Coton (5) cited discrepancies in the above method when blade experiences very high angles of attack, even though 3-D effects were included. He also cited that the predictions overshoot the impulsive response of the blade at the exit from the tower shadow region.

The two schools of thought both suggest that 3-D effects cannot be neglected to determine the effect of blockage on rotating blades. It is important that further studies include three dimensional effects. Raghav (6) cites that the nature of the separation line on the rotating blade must involve interaction between the stall vortex and the details of separated flowfield downstream. Thus, study of this region opens the way to understanding 3-D flow separation leading to Dynamic Stall, Raghav (7). The motivation of the present studies is two-fold. The first is to study the effect of blockage on a rotating blade. Second is to study the significance of radial flow during transient or dynamic stall.

In this work the ñinflowò onto a rotating blade is obstructed to study the characteristics of the flow over the blade under the imposed conditions. The obstruction here is an extreme case of the practically occurring cases of the wind turbines and radiator fans. The flow behind the stall line is of interest as the radial acceleration is strongest in that region. In practice a rotating blade interacting with a blocked flow occurs in the case of downwind Horizontal Axis Wind Turbines (HAWT), when the blade passes through the tower ñshadowò. Such stall also occurs on automobile radiator fan blades, where the flow is blocked by the other parts. This phenomenon gives rise to dynamic loading on the blades which is hard to predict. In case of HAWT it leads to increased fatigue loading leading to structural failure and lower life expectancy. The dynamic loading on a radiator fan leads to increased noise generation by the blades of the fan.

2. PREVIOUS WORK

The effect of blockage over a part of the azimuth of a rotating blade is of interest in many practical applications. It is known that flow separation leading to stall, occurs on a rotating blade interacting with a blocked flow. This flow separation is dynamic in nature due to which the reattachment is delayed in a hysteresis loop

which persist well into the next quadrant of the rotation. Butterfield (8) cites that dynamic stall has been observed on wind turbine blades due to the effect of the tower shadow. Efforts to capture this phenomenon have proceeded through 2-D experiments on pitching airfoils in wind and water tunnels McCroskey (1) (9). Numerical predictions proceeding from inviscid formulations to full Navier-Stokes simulations have also been performed. Research on 3-D flow separation leading to dynamic stall showed that the stall initiation was a localized event, thus making the timing of stall difficult to predict and control.

Influence of Radial Acceleration

The effect of radial acceleration on the lift and pitching moment evolution on a dynamically stalled blade is of great interest. Coton (10) cites the substantial pressure differences in the inboard pressure distribution on rotating wind turbine blades compared to 2-D models, preceding dynamic stall. He also cites the delay in forwards movement of the separation region (prior to lift-off of the dynamic lift vortex) due to spanwise flow. Corten (11) has used a ñstall flagò method to capture the occurrence of stall on full scale wind turbine blades. His analysis shows that in dynamic stall region, there is a substantial radial pressure gradient and accompanying radial flow. Corten's (12) laser velocimeter results shows the formation of vorticity directed parallel to the rotor axis along the aft portions of the blade. Hence radial accelerations must have a strong influence on the lift and pitching moment evolution after stall occurs. Xu (13) shows that the influence of rotation (centrifugal effects in the boundary layer) on a wind turbine blade, when modeled using a full Navier-Stokes formulation, shows a substantial stall delay in the 3-D case compared to 2-D cases where there is no rotation effect.

3. MEASUREMENT APPROACH

The objective of this study is to study the flow field behind the point of separation on inboard region of a rotating blade under blockage effects. The approach was to use a single-bladed rotor in a hover facility and considerably alter its effective angle of attack over a part of the azimuth. The sudden change in angle of attack was achieved by using a ñflow obstructerò plate in the inflow region of the rotor. The primary concern is to determine the effect of blockage on a rotating blade. The secondary concern is the behavior of radial flow downstream of the separation line on the blade.

Experimental Facility

The experiment was conducted in a rotorcraft hover facility in the School of Aerospace Engineering at Georgia Institute of Technology. A single bladed rotor of the specifications shown in Table 1 was used for the experiments. The Figure 1 shows a picture of the experimental setup and a schematic front view of the facility. Refer to Yang (14) for the complete details of the experiment.

Description	Value	Units
Blade Radius	0.61	Meters
Blade Chord	0.13	Meters
Aspect Ratio	4.8	
Airfoil Section	NACA 0012	
Blade	Un-Tapered, Untwisted	
Motor	11.2	KW
Motor RPM	0-200	RPM

Table 1 Hover Facility Specifications

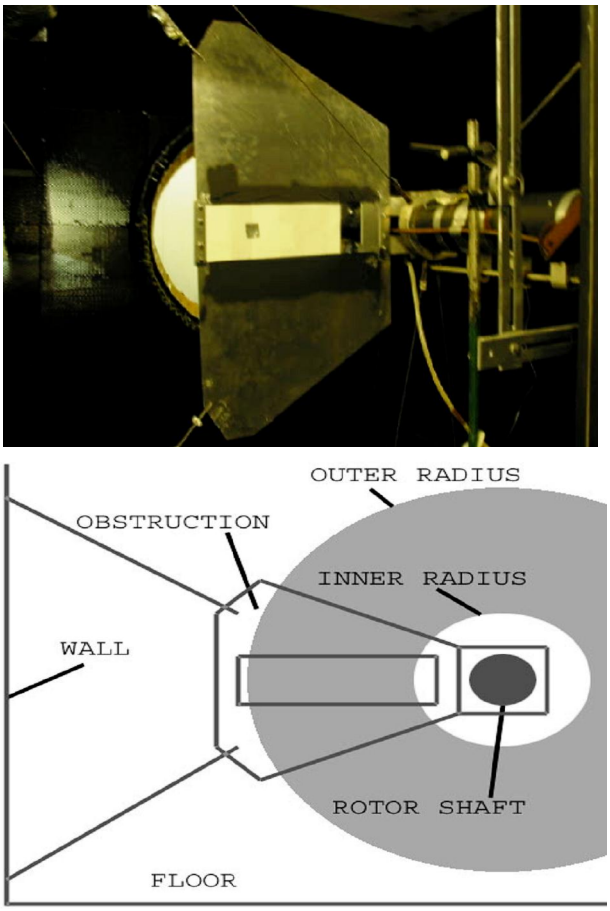


Figure 1: Inflow obstruction picture and sketch of inflow region

4. FLOW MEASUREMENTS

From prior experiments in this facility, Liou (15) , it is known that the blade used here operates with attached flow at 15 degrees pitch, at 500 RPM (30 m/s tip speed).

Qualitative Visualization

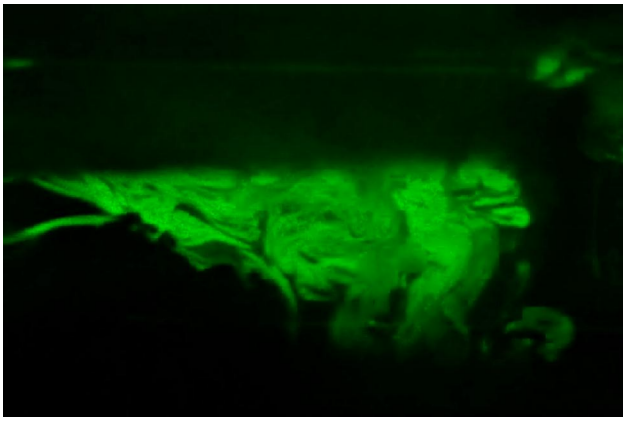
Pulsed Laser Sheet Imaging (PLSI) studies were conducted at 70 RPM. The RPM was slow enough to capture successive video images, showing the same seeding during a single-event approach and passage of the blade through the seeding cloud, albeit at low Reynolds number. The PLSI showed that the flow across the disc (axial velocity component), was being stopped suddenly in contrast to being sharply accelerated as a lifting blade passes by, this indicated blade stall. In Figure 3, as the blade moves through the seeding cloud, very little axial flow movement can be seen (a) and b)), showing the absence of strong suction. Soon after the blade passes through, the seeding resumes its downward movement. (c). Viewed in detailed image sequences, this visualization made clear that the blade was essentially just a blunt body moving through the flowfield, and was not generating any significant lift. The inflow during the rest of the cycle was due to the lift generated when the blade was beyond the inflow-obstruction region. This finding is quantified by PIV.

Quantitative Visualization

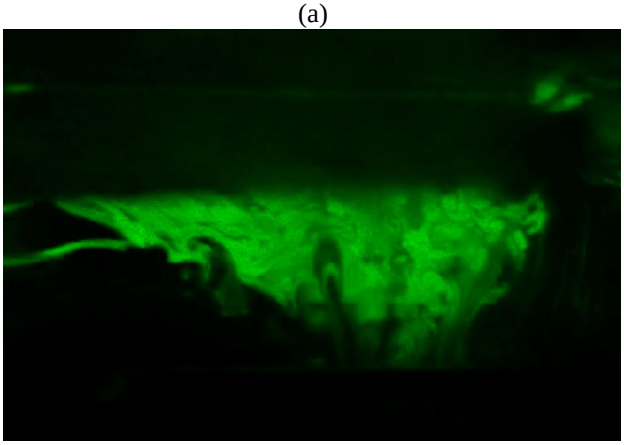
For PIV, four image sets of fifty image pairs were taken for each blade azimuth of the measuring plane. The spanwiseï axial section was selected to capture the radial flow velocity (Figure 4). The measuring plane was illuminated along the span with a rectangular viewing region of 12.8 cm (spanwise) x 9.6 cm (axial).

5. EVIDENCE OF STALL

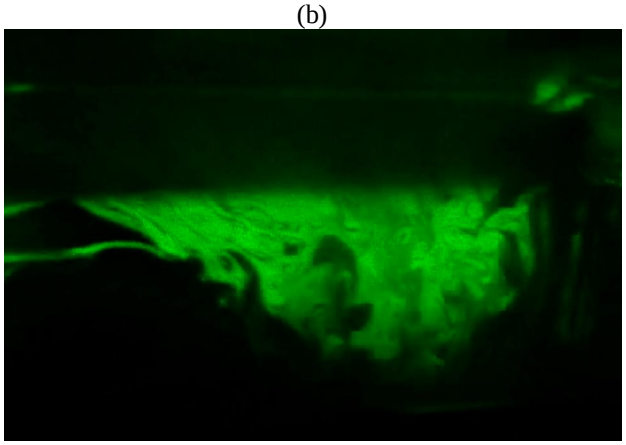
The experiment was conducted under conditions where the aerodynamic load is small compared to the inertial loads. Also, no direct thrust measurement is made, and no pressure sensors are used. Thus, the occurrence of stall was verified by comparing the axial velocity induced by blade passage with the ñlocalò level attributable to the tip vortex system and lift generation through the rest of the blade cycle. Previous measurements with a lift-generating rotor have shown that a very large inflow velocity spike occurs during blade passage, followed by settling to the ñlocalò level very quickly for the rest of the cycle. However, if the blade is stalled as it passes the measuring location, and generating lift through much of the rest of the cycle, we should expect to see only the smaller perturbation in the inflow velocity during blade passage, associated with passage of a blunt body. It is expected that the flow will be pushed away by thickness of the blade, followed by a briefly increased inflow. Velocity results from the 70 RPM tests, followed by cleaner results at 500 RPM, verified that this was occurring. Figure 4 shows that at mid-span, at 270 degrees azimuth (blade is at the opposite side of the rotor disc) the measured inflow velocity was about 2.5 m/s for the obstructed case and 3 m/s for the control case, as shown in Figure 4. This is in line with rough expectations from Momentum Theory.



(a)



(b)



(c)

Figure 3 Sequential Blade and Seeding Visualization and Schematic View

6. FLOW CHARACTERISTICS – RADIAL FLOW

Figures 5 through 8 show contours of the radial velocity component, measured at various locations along the spanwise viewing plane covering 12.8 cm of the blade centered on the 38 cm radial location (midspan). At the leading edge plane, (Figure 5), without the obstruction, at 90 degrees azimuth, when the blade is assumed to be lifting and at the measurement region, the radial velocity component is in the range of 0.5 m/s (red region).

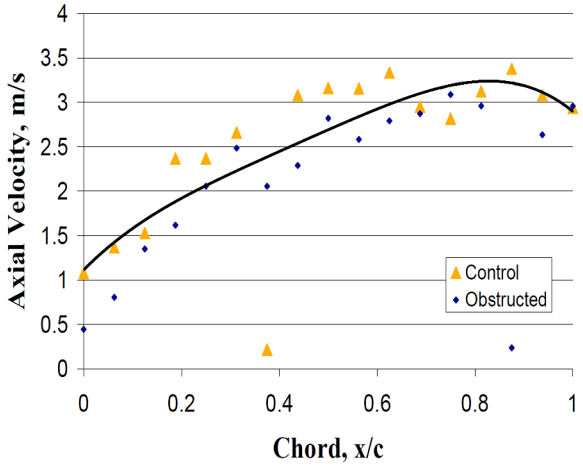


Figure 4 Axial Velocity Profile at Mid-span

However, when the blade is at 270 degrees (as far away as possible from the measuring region) the radial velocity is on the order of 1.4 m/s, as shown in Figure 6. This is the ñlocalò level of radial velocity, attributed to the tip vortex.

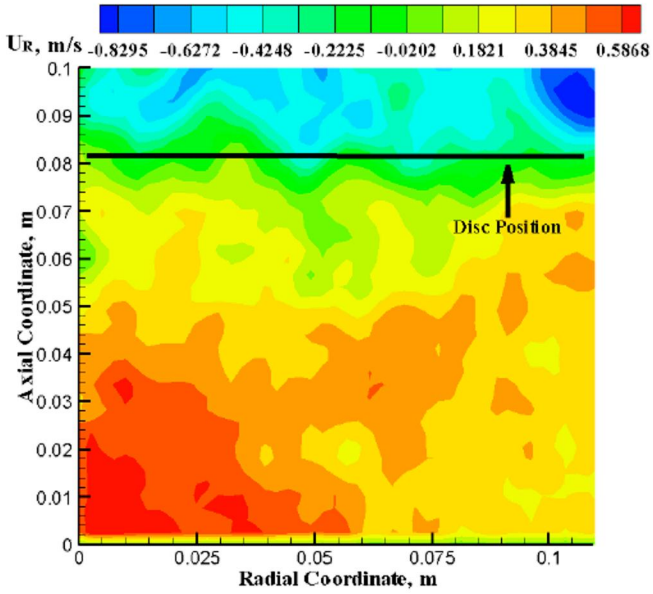


Figure 5 Radial Velocity Contour at Leading edge, No Obstruction

Comparing Figures 7 and 8 for the same cases as above, but with the obstruction present, we see that the radial velocity remains near the local level. Thus in the case with obstruction, the radial velocity above the blade leading edge region is not very different from the local. This sets the context to view the change of radial velocity from leading edge to trailing edge over the blade, as it Moves through the measurement plane, and compare the cases with and without obstruction.

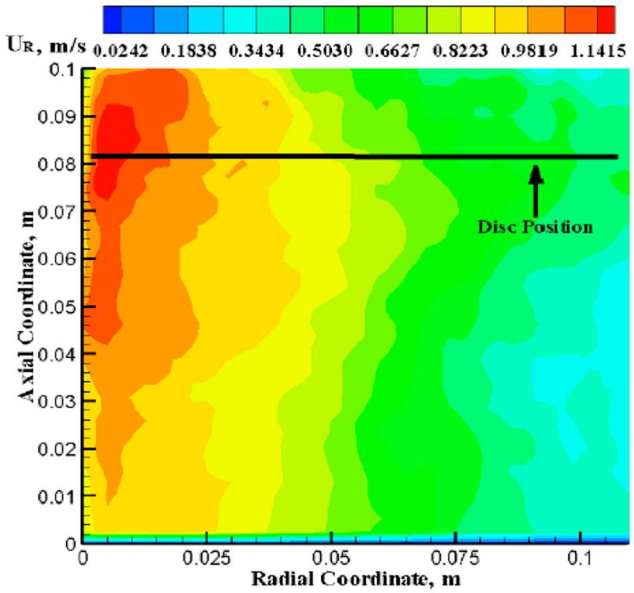


Figure 6 Radial Velocity Contour with blade at 270 degrees azimuth, no obstruction

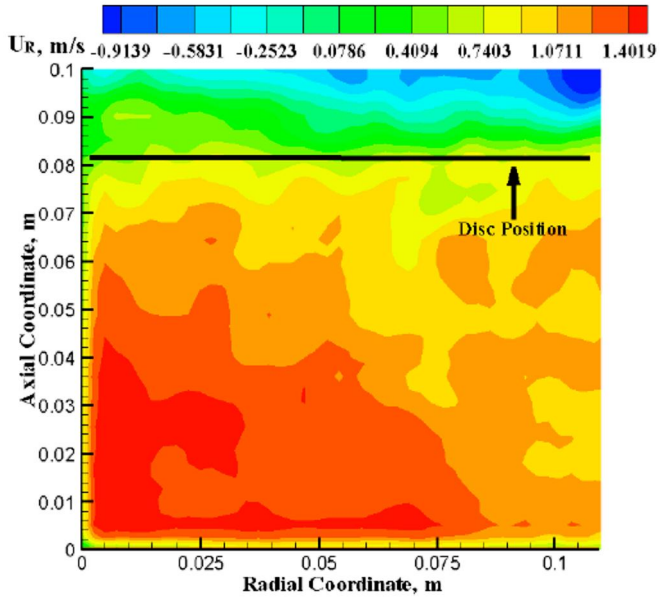


Figure 7 Radial velocity contours at leading edge, obstructed case.

The measurements at mid-span are shown. In Figures 9 and 10, the blue lines show the deceleration from leading edge to quarter chord at 5 cm from the surface of blade, while the orange line is 10 cm from the surface. Figure 10 shows the radial velocity as a function of chord at mid-span without the obstruction. There is less discontinuity after the quarter chord than in the case with obstruction (Figure 10).

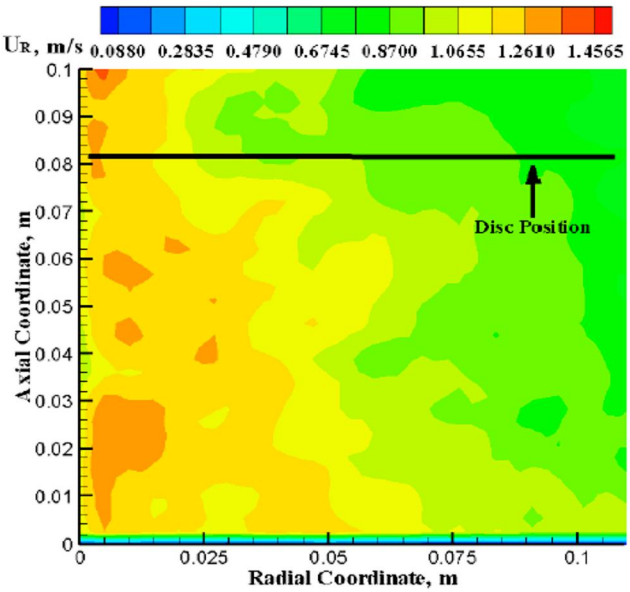


Figure 8 Radial velocity contours with blade at 270 degrees azimuth, obstructed case

The case with the obstruction, in Figure 10, shows a quick change in the magnitude of the radial velocity. The velocity fluctuation increases after the quarter chord. The reason why the radial inflow velocity is initially higher in the obstructed case is that while the axial flow has been obstructed, the radial flow has not been obstructed. So when the pressure drop occurs near the upper surface, the radial inflow is accelerated as opposed to the controlled case where axial component of the flow would be sufficient in pressure recovery.

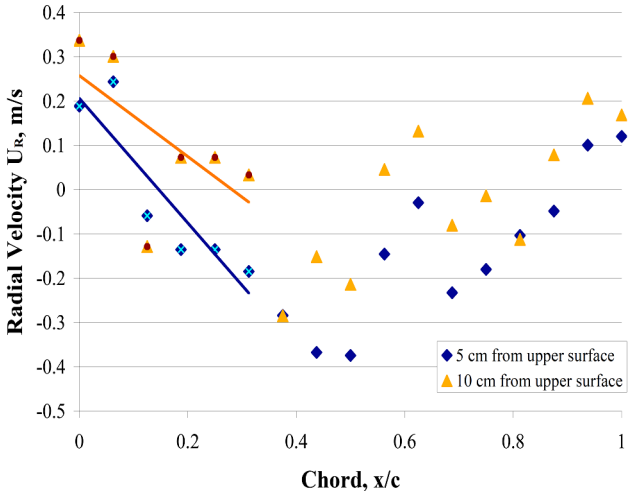


Figure 9 Radial Velocity Profile at Mid-Span without obstruction (Control Case)

The range of the radial inflow velocity variation for both cases (controlled and obstructed) is the same up to the quarter chord, where flow is believed to be attached for both cases. However the rate of deceleration of radial inflow in the control case decreased over the observed distance from the blade surface.

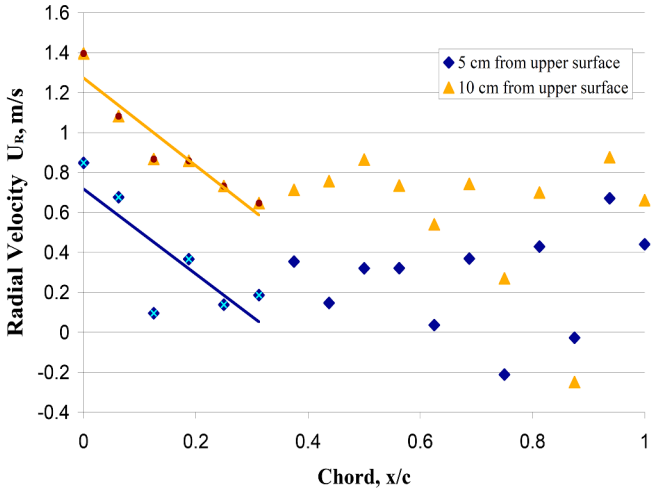


Figure 10 Radial Velocity Profile at Mid-Span with Obstruction

In contrast, the obstructed case had the same deceleration at 5 cm and 10 cm away from the surface. The results show that the inboard-directed radial component of the inflow velocity decreases as the blade passes through the obstruction. This suggests that the centrifugal effect dominates, downstream of quarter chord. After the blade has passed through, the radial velocity returns to the local level.

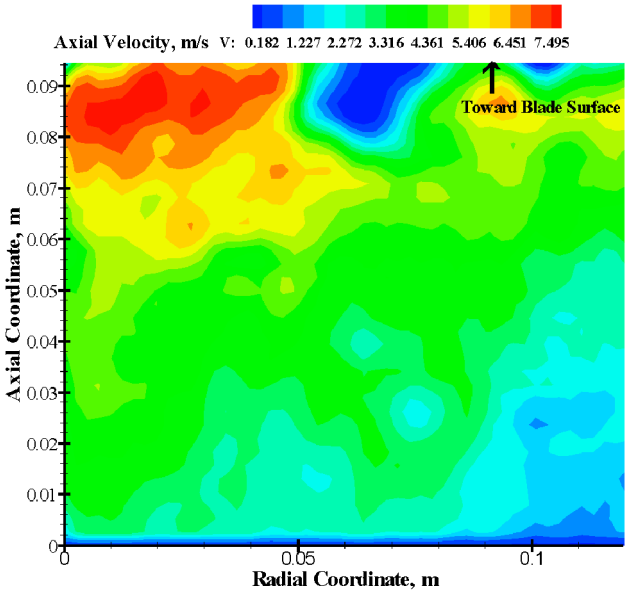


Figure 11 Instance of Recirculation at 0.25 x/c

7. AXIAL FLOW

Figures 11 and 12 show two instances of the span-wise cells in the separated flow. This process does not appear to be periodic at the rotor frequency, and thus differs substantially in location and strength from one rotor cycle to another. Thus this feature does not show up in averaged measurements. However examination of several PIV frames confirms that the phenomenon persists.

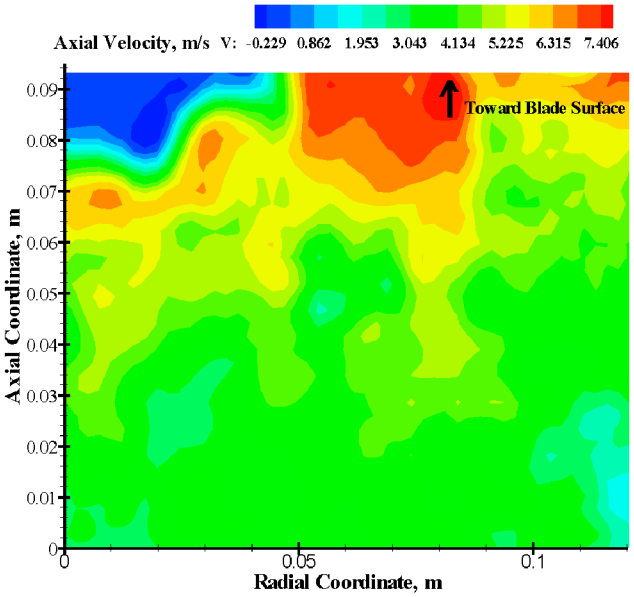


Figure 12 Instance of Recirculation at 0.28 x/c

8. CONCLUSIONS

A single-blade hover facility is used as a test-bed to study the temporal evolution of obstructed flow phenomena close to a rotating blade. The transient stall phenomenon is studied, given that the interest is in the region downstream of the separation line.

1. PLIS and PIV validate the experiment in that stall indeed occurs, based on the blockage of the inflow.
2. Downstream of the stall line, the radial velocity along the blade sharply develops an outward direction. However, the formation of stall cells along the radius serves to exchange fluid away from the blade boundary layer.
3. The phase-locked radial velocity profile shows higher deceleration of radial out-flow from leading edge to chord-wise location where flow recirculation begins. This suggests that while reattached flowôs inboard velocity component started to dominate over the centrifugal effect from leading edge to separation point, the inflow directional preference is nullified by the separation/recirculation zone created.
4. The cells of radial flow separation, while in the comparable scale as the blade chord, is time varying. This is possibly due to the dynamic nature of the extent of the radial out-flow due to domination of the centrifugal effect.

REFERENCES

1. **McCroskey, W.J., McAlister, K.W., Carr, L.W., and Pucci, S.L.** *An Experimental Study of Dynamic Stall on Advanced Airfoil Sections.* s.l. : NASA-TM, 1982.

2. **Chadrashekara, M.S., Wilder, M.C., Carr, L.W.** *Compressible Dynamic Stall Control Using Dynamic Shape Adaptation.* : AIAA, 1999. 99-0655.

3. **McAlister, K. W. and Tung, C.** *Suppression of Dynamic Stall with a Leading-Edge Slat on a VR- 7 Airfoil.*NASA, March 1993. TP 3357.

4. **Leishman, J.G.** *Challenges in Modelling the Unsteady Aerodynamics of Wind Turbines.* Wiley - Wind Energy, 2002.

5. **Tongguang Wang, Frank N. Coton.** *A High Resolution Tower Shadow model for Downwind Wind Turbines.* Journal of Wind Engineering And Industrial Aerodynamics - Elsevier, 2001.

6. **Raghav, V., Richards, P., Smith,M., Komerath,N.** *Three-Dimensional Features of the Stalled Flow Field of a Rotor Blade in Forward Flight.* Seoul Oct 2009.

7. **Raghav, V., Richards, P., Smith,M., Komerath,N.** *An Exploration of the Physics of Dynamic Stall.* AHS, 2009.

8. **Hansen, A.C., Butterfield, C.P** *Aerodynamics of Horizontal-Axis Wind turbines..*Annual Review of Fluid Mechanics, 1993, Vol. 25.

9. **McCroskey, W. J** *Some Current Research in Unsteady Fluid Dynamics - The 1976 Freeman Scholar.* Journal of Fluids Engineering, 1977, Vol. 99.

10. **Coton, F. N., Wang, T., and Galbraith, R. A. McD.** *An Examination of Key Aerodynamic Modelling Issues Raised by the NREL Blind Comparison.* Wind Energy, 2002. 1095-4244.

11. **Corten, Gustave.P.,.** *Flow Separation on Wind Turbine Blades.* University of Utrecht : Doctoral Dissertation, 2001.

12. **Bunnes, P., Fiddes, S.,.** *Laser Doppler Results Contributed to 'Dynamic Stall and Three Dimensional Effects.* Glasgow, UK Joule II contractors meeting, 1995.

13. **Xu, Guanpeng.** *Computational Studies of Horizontal Axis Wind Turbines.* PhD Thesis, Georgia Instiute of Technology, 2001.

14. **Jaesuk Yang, Balakrishnan Ganesh, Narayanan Komerath.** *Radial Flow Measurements Downstream of Forced Dynamic Separation on a Rotor Blade.* San Francisco, California AIAA Fluid Dynamics Conference and Exhibit, 2006. 2006-3377.

15. **Liou, S-G., Hyun, J-S.,Komerath, N.M.** *Flowfield of a Swept Blade Tip at High Pitch Angles* AIAA, 1991. 914-0704.

16. **Bousman, W.** *A Quantitative Examination of Dynamic Stall from Flight Test Data.* AHS Forum, 1997.

17. **Bousman, W.** *Evaluation Of Airfoil Dynamic Stall Characteristics For Maneuverability.* European Rotorcraft Forum, 2000.

OPTIMIZATION OF OPERATIONAL CONDITIONS TO THE TREATMENT OF VINASSES USING ENTRRAPED ACTIVATED CARBON: AN INCOMPLETE FACTORIAL DESIGN

José Manuel CRUZ
Chemical Engineering Department, University of Vigo
Vigo, 36310 / Campus As Lagoas, Spain

Xanel VECINO
Chemical Engineering Department, University of Vigo
Vigo, 36310 / Campus As Lagoas, Spain

Rosa DEVESA-REY
Chemical Engineering Department, University of Vigo
Vigo, 36310 / Campus As Lagoas, Spain

and
Ana Belén MOLDES
Chemical Engineering Department, University of Vigo
Vigo, 36310 / Campus As Lagoas, Spain

ABSTRACT

In this study, entrapped activated carbon was proposed to remove colored and phenolic compounds from vinasses. Operational conditions of batch adsorption process were optimized by means of an incomplete 3^3 factorial design. The independent variables studied consisted on the volume of entrapped activated carbon (x_1); time of treatment (x_2) and agitation speed (x_3); whereas the dependent variables studied were the absorbance reduction measured at 520 nm (y_1) and the absorbance reduction measured at 280 nm (y_2). The optima conditions to remove red colored compounds and phenolic compounds in batch adsorption experiments, from vinasses, consisted on entrapped activated carbon/vinasse ratio about 1.5; operational time=1.5 h., and agitation speed about 100 rpm. So, the results obtained in this study evidenced the potential application of immobilized activated carbon beads in the design of new eco-friendly, efficient and low-cost bioremediation strategies.

Keywords: *entrapped activated carbon, alginate beads, phenolic compounds, red colored compounds, incomplete experimental design*

1. INTRODUCTION

Industrial activities, agriculture and the elimination of landfill wastes have led to contamination of aquifers. Vinasses are colored wastewaters from wine industry that could contaminate these aquifers.

Activated carbon is an adsorbent widely used in industrial processes because of its ability to adsorb large amounts of substances of different nature, whether organic compounds, heavy metals and colored substances in nature that create a significant environmental impact when they are discharged to the environment. The usual activated carbon marketing is in granular or powder form, although the latter presents a more effective adsorption processes due to their high surface area. However the activated carbon presents a series of disadvantages that stem mainly from their high surface area. These disadvantages are its difficult handling, as well as its ability to clog membranes. To avoid such disadvantages, an immobilization of activated carbon into calcium-alginate beads may be carried out, thereby encouraging its use, regeneration, and storage. Alginate salts, especially calcium alginate, are frequently used to immobilize microorganisms or enzymes. Other applications of calcium alginate are based on the immobilization of compounds with adsorption capacity such as coal, chitosan [1], polyvinyl alcohol [2], carboxylates [3], and humic acids[4].

Calcium-alginate beads have been employed for remediation purposes in ground water contaminated areas by designing permeable reactive barriers (PRBs) [5, 6].

The main objective of this study is to evaluate the efficiency of entrapped activated carbon in the decontamination of colored wastewaters from wine industries, during batch adsorption process. For this purpose an incomplete factorial design was applied. The independent variables studied consisted on the ratio volume of entrapped activated carbon/volume of vinasses; time of treatment and agitation speed; whereas the dependent variables studied were the absorbance reduction measured at 520 nm and the absorbance reduction measured at 280 nm.

2. MATERIALS AND METHODS

Immobilization of activated carbon

Entrapped activated carbon was prepared according to the procedure described by Devesa-Rey et al. [7]. Sodium alginate (4%), was mixed with activated carbon (2%). Following the mixture was added drop wise from a 1.5 ml pipette to a calcium chloride solution (0.58 M), which was employed as crosslinking solution, by employing a constant pump (Masterflex L/S). **Figure 1** shows a picture of the entrapped activated carbon employed in this work.

Figure 1. Picture of activated carbon entrapped in alginate beads.

Box-Behnken experimental design

Response surface methodology consists of a group of mathematical and statistical techniques based on the fit of empirical models to the experimental data obtained in relation to experimental design [8]. Box–Behnken designs are a class of rotatable or nearly rotatable second-order designs based on three-level, incomplete factorial designs [9]. The number of experiments (N) required for full Box-Behnken design are given by the formula $N = 2k\ (k-1) + C_0$, where k is the number of factors and C_0 is the number of central points [10]. The simplest equation describing a linear function is described by equation 1.

$$y = \beta_0 \sum_{i=1}^{k} \beta_i x_i + \varepsilon \qquad eq.\ (1)$$

Where β_0 is the constant factor; β_i represents the coefficients of the linear parameters; k is the number of variables; x_i represents the variables; and ε is the residual factor associated to the experiments. When the experimental data are not fitted by linear equation, then it is desirable to include levels in the input variables. In this case, a polynomial response surface must be generated. The Box-Behnken experimental designs were constructed for situations in which it was desirable to fit a second-order model (equation 2).

$$y = \beta_0 \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \sum_{j \geq i}^{k} \beta_{ij} x_i x_j + \varepsilon \qquad eq.\ (2)$$

The coefficients β_{ij} represent the interaction parameters. These designs include a central point employed to determine the curvature, and the determination of critical or optima conditions are deduced from the above second-order function by including quadratic terms (equation 3).

$$y = \beta_0 \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k} \sum_{j \geq i}^{k} \beta_{ij} x_i x_j + \varepsilon \qquad eq.\ (3)$$

The coefficients β_{ii} represent the quadratic parameters. So, the experimental data allow the development of empirical models describing the interrelationship between operational and experimental variables by equations including linear, interaction and quadratic terms.

The range of independent variables employed in this work is included in **Table 1,** as well as the dependent variables studied.

Table 1. Independent and dependent variables employed in this study.

a) Independent variables			
Variable	**Nomenclature**	**Units**	**Variation range**
Ratio entrapped activated carbon / vinasses	ratio	v/v	0.5 – 1.5
Time	t	h	0.5 – 2
Speed agitation	speed	rpm	75 – 150
b) Dimensionless, coded independent variables			
Variable	**Nomenclature**	**Definition**	**Variation range**
Dimensionless ratio entrapped activated carbon/ vinasses	x_1	(ratio – 1)/0.5	(–1,1)
Dimensionless time	x_2	(t – 1.25)/0.75	(–1,1)
Dimensionless speed agitation	x_3	(speed – 112.5)/37.5	(–1,1)
c) Dependent variables			
Variable	**Nomenclature**	**Units**	
Absorbance reduction at 520 nm	y_1	%	
Absorbance reduction at 280 nm	y_2	%	

So, the quadratic function obtained for the three variables is described in equation 4.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2$$

eq. (4)

Where y is the dependent variable, β denotes the regression coefficients (calculated from experimental data by multiple regressions using the least-squares method) and x denotes the independent variables. The experimental data were analyzed by the Response Surface method using the Statistica 7.0 software.

Batch adsorption experiments

Adsorption experiments were carried out in 250 mL Erlenmeyer flasks, using different ratios entrapped activated carbon/vinasses and different agitation speed (see **Table 1**).

3. RESULTS AND DISCUSSION

The standardized (coded) dimensionless independent variables employed, with variation limits (-1, 1), were defined as x_1 (ratio volume of entrapped activated carbon/volume of vinasses), x_2 (operational time) and x_3 (agitation speed). The correspondence between coded and uncoded variables was established by linear equations deduced from their respective variation limits, according to equation 5 [8]:

$$x_i = \left(\frac{z_i - z_i^0}{\Delta z_i} \right) \beta_d \qquad eq.(5)$$

Where Δz_i is the distance between the real value in the central point and the real value in the superior or inferior level of a variable; β_d is the major coded limit value in the matrix for each variable; and z^0 is the real value in the central point. Coded variables are then assigned values of -1, 0 and +1, corresponding to the lowest, central and maximum variation limits for each variable. Thus, the response surface obtained from the coded variables is not influenced by the magnitude of each variable, allowing the combination of factors into a dimensionless scale.

Table 2 shows the set of experimental conditions assayed (expressed as coded variables) and the experimental data obtained, for variables y_1 and y_2.

Besides, **Figure 2** and **Figure 3** show the most influential variables on the percentage of absorbance reduction at 520 and 280 nm respectively. It can be observed that the most influential variables for the absorbance reduction at 520 nm was the adsorption time followed by the ratio volume of entrapped activated carbon/ volume of vinasses; whereas on the percentage of absorbance reduction at 280 nm the most influential variable was the ratio activated carbon/vinsasses followed by the adsorption time. For both dependent variables the agitation speed was the less influential independent variable.

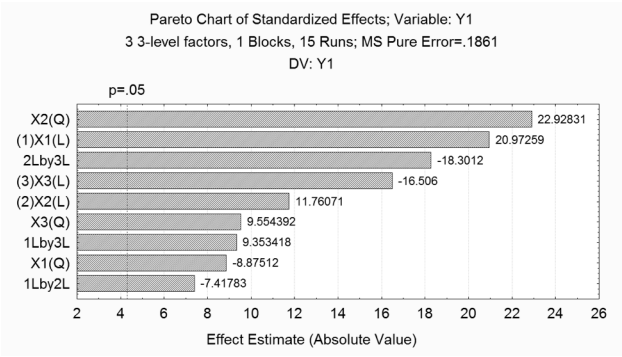


Figure 2. Pareto chart of the standardized effects on the absorbance at 520 nm.

Table 2. Operational conditions considered in this study (expressed in terms of the coded independent variables) and experimental results achieved for the dependent variables assayed, y_1 (absorbance reduction at 520 nm), and y_2 (absorbance reduction at 280 nm).

Independent Variables					
Exp.	x_1	x_2	x_3	y_1	y_2
1	0	–1	–1	77,24	91,35
2	0	1	–1	86,06	94,62
3	0	–1	1	71,73	91,70
4	0	1	1	64,76	93,94
5	–1	–1	0	72,76	87,19
6	–1	1	0	82,21	93,80
7	1	–1	0	79,16	93,08
8	1	1	0	82,21	94,81
9	–1	0	–1	77,64	88,54
10	–1	0	1	76,94	89,70
11	1	0	–1	83,20	94,55
12	1	0	1	90,57	96,57
13	0	0	0	81,79	94,60
14	0	0	0	82,65	94,75
15	0	0	0	82,28	94,72

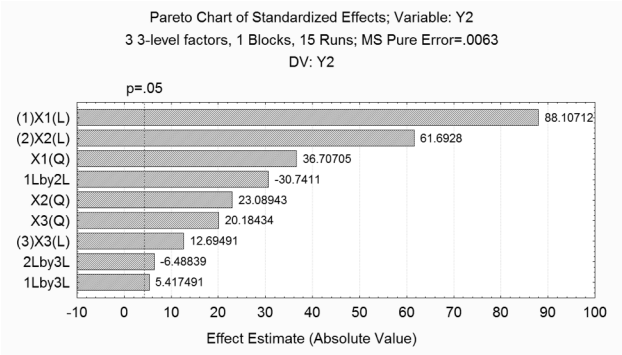


Figure 3. Pareto chart of the standardized effects on the absorbance at 280 nm

Additionally, **Table 3** shows the regression coefficients and their statistical significance for variables y_1 and y_2. It can be observed that all the coefficients are statistically significant at $p<0.05$.

Table 3. Regression coefficients and their statistical significance for variable y_1 and y_2, corresponding to percentage of absorbance reduction at 520 nm and 280 nm respectively.

Regression Coefficients	y_1	p_{y1}	y_2	p_{y2}
b_0	82,24	9,12467E-06	94,69	2,64963E-07
b_1	3,20	0,002255994	2,47	0,000145827
b_{11}	1,99	0,012403422	-1,52	0,000837775
b_2	1,79	0,007113946	1,73	0,000296941
b_{22}	-5,15	0,001887184	-0,95	0,002111993
b_3	-2,519	0,003626745	0,36	0,006904164
b_{33}	-2,15	0,010725757	-0,83	0,002779761
b_{12}	-1,60	0,017644445	-1,22	0,001198369
b_{13}	2,02	0,011169744	0,21	0,036490652
b_{23}	-3,95	0,002957403	-0,26	0,026131355

On the other hand, **Figure 4** and **Figure 5** show the variation of absorbance reduction at 520 and 280 respectively with the most influential variables (x_1) and (x_2); fixing the variable x_3 at intermediated values. The maximum percentage reduction of absorbance at 520 nm was about 87 % whereas the maximum percentage of absorbance reduction at 280 nm was about 95 %.

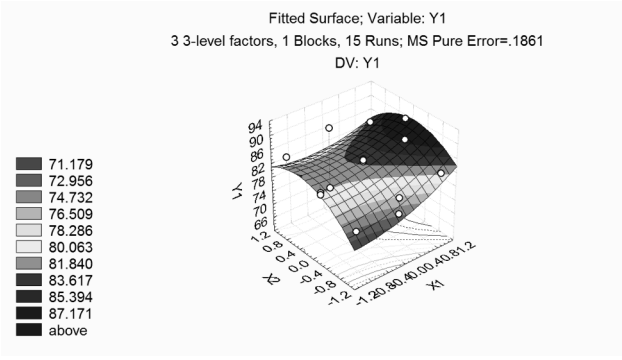


Figure 4. Variation of absorbance reduction at 520 with the most influential variables (x_1=ratio entrapped activated carbon/vinasses) and (x_2= operation time); fixing the variable x_3 at intermediated values.

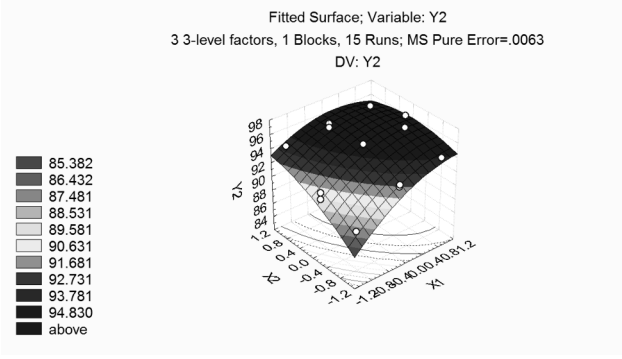


Figure 5. Variation of absorbance reduction at 520 with the most influential variables (x_1=ratio entrapped activated carbon/vinasses) and (x_2= operation time); fixing the variable x_3 at intermediated values.

Besides **Figure 6** and **Figure 7** show observed versus predicted values for the variables y_1 and y_2 respectively, which that accounts the fit of the model between predicted and observed values.

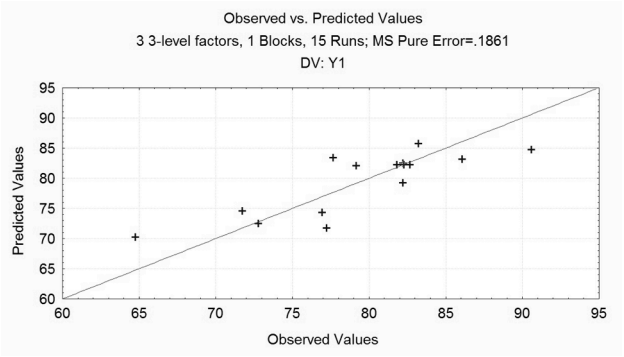


Figure 6. Observed versus predicted values for the absorbance reduction at 520 nm.

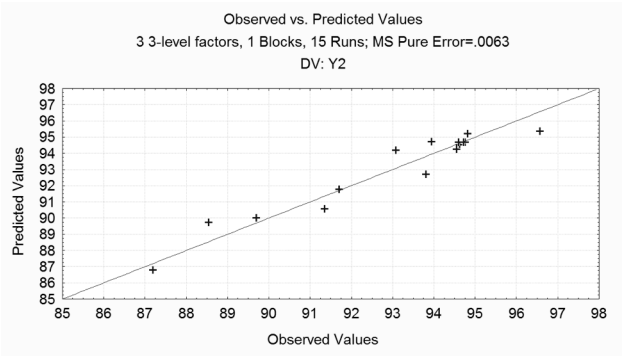


Figure 7. Observed versus predicted values for the absorbance reduction at 280 nm.

4. CONCLUSIONS

Immobilized activated carbon beads proven to be efficient in the removal of red color from vinasse. From the above results it can be concluded that the maxima percentages of absorbance reduction at 520 nm and 280 nm can be achieved using the highest activated carbon/vinasse ratio (1.5) at intermediate operational time (1.5 h), and intermediate agitation speed (112 rpm).

Acknowledgements

The authors wish to thank to the Xunta de Galicia Regional Government for financial support (project Ref. 2009/070) and to the Ángeles Alvariño Program.

5. REFERENCES

[1] T. Gotoh, K. Matsushima and K.I. Kikuchi, "Preparation of alginate-chitosan hybrid gel beads and adsorption of divalent metal ions", **Chemosphere**, Vol. 55, Nº 1, 2004, pp. 135-140.

[2] C. Jeon, J.Y Park and Y.J. Yoo, "Novel immobilization of alginic acid for heavy metal removal", **Biochemical Engineering Journal**, Vol. 11, No. 2-3, 2002, pp. 159-166.

[3] C. Jeon, J.Y Park and Y.J. Yoo, "Characteristics of metal removal using carboxylated alginic acid", **Water Research**, Vol. 36, No. 7, 2002, pp. 1814-1824.

[4] A.K. Pandey, S.D. Pandey, V. Misra and S. Devi, "Role of humic acid entrapped calcium alginate beads in removal of heavy metals", **Journal of Hazardous Materials**, Vol. 98, No 1-3, 2003, pp. 177-181.

[5] S. Krajangpan, J.J. Elorza Bermudez, A.N. Bezbaruah, B.J. Chisholm and E. Khan, "Nitrate removal by entrapped zero-valent iron nanoparticles in calcium alginate", **Water Science and Technology**, Vol. 58, No. 11, 2008, pp. 2215-2222.

[6] X. Liu, L. Zhao, S. Liu and C. Cui, "Study for remediation of groundwater contaminated by BTEX using a biological barrier", **Advanced Materials Research**, Vol. 178, 2011, pp. 254-259.

[7] R. Devesa-Rey, G. Bustos, J.M. Cruz and A.B. Moldes, "Optimization of entrapped activated carbon conditions to remove colored compounds from winery wastewaters", **Bioresource Technology**, (doi:10.1016/j.biortech.2011.03.072) 2011, *in press*.

[8] M.A. Bezerra, R.E. Santelli, E.P. Oliveira, L.S. Villar and L.A. Escaleira, "Response surface methodology (RSM) as a tool for optimization in analytical chemistry", **Talanta**, Vol.76, 2008, pp. 965–977.

[9] G.E.P. Box and D.W. Behnken, "Simplex-sum designs – A class of 2nd order rotatable designs derivable of those of 1st order". **An Math Statistics,** Vol. 31, No. 4, 1960, pp. 838–864.

[10] S.L.C. Ferreira, R.E. Bruns, H.S. Ferreira, G.D. Matos, J.M. David, G.C. Brandão, E.G.P. da Silva, L.A. Portugal, P.S. dos Reis, A.S. Souza and W.N.L. dos Santos, "Box-Behnken design: An alternative for the optimization of analytical methods", **Analytical Chimica Acta**, Vol. 597,No. 2, 2007, pp. 179–186.

An Exploration of Crosscutting Concerns in Software Requirements and Design

Jorge Fox*
Department of Telematics
Norwegian University of Science and Technology
NO-7491, Trondheim, Norway
jfox@item.ntnu.no

ABSTRACT

Aspect orientation is a software engineering technique that provides an enhanced separation of concerns. The idea is that concerns that affect several modularization entities, that is crosscutting concerns, (C^3) can be better managed by first identifying and then weaving them into selected entities. This is supposed to improve modularization of software, facilitate reuse of code or software components, and support software evolution. In this work, we explore C^3 and their relation to aspects. The thesis is that crosscutting originates during the transformation of requirements into software design or specification entities. We define C^3 through a tracing relation from requirements into specifications and from specifications into design entities. We illustrate the translation of requirements into use cases and these into formal specifications in a first order language, Alloy. These relations help to identify candidate C^3 from the specifications. Furthermore, we propose a classification of C^3 based on the classification of requirements.

Keywords

Software Development, Aspect Orientation, Software Abstractions, Requirements Engineering

1. INTRODUCTION

Aspect Orientation (AO) aims to provide one more dimension for modularisation of software. The idea is that computer systems are better programmed by separately specifying the various concerns, understood as properties or areas of interest of a system. Concerns are described separately and then weaved, that is, are composed relying on mechanisms in the underlying aspect oriented environment. Making the final product a coherent software system.

This description of AO is uncomplicated, though it conceals a number of problems.

In this paper, we explore the source of aspects in the tracing from requirements to the subsequent phases in software development, particularly from informal requirements to formal specifications. While reflecting upon the question "How can we define aspects beyond the programming level?" we came to the conclusion that crosscutting can be defined by the transformation from one abstraction level to another. We explore the source of crosscutting at the hand of a case study and illustrate the translation of requirements to specifications. We consider requirements as belonging to the problem space of the system. These requirements are translated into a system model based on a selected design framework such as classes in object orientation. In our work we further specify the requirements in a next step in the lightweight formal language Alloy. We show that C^3 are specifications defined over several modules or classes. This way we provide C^3 with meaning in terms of requirements tracing by defining them as entities that improve the base modularisation paradigm, that is, entities that capture or implement C^3. Crosscutting is one of the defining characteristics of aspects. Moreover, we argue that there are two main kinds of crosscutting and classify aspects into two broad groups: functional and non functional.

The rest of the paper is structured as follows. In Sect. 2 we introduce the definitions used in our framework. In Sect. 3 we describe a transformation example of requirements into specifications. In Sect. 4 we identify the sources of crosscutting according to our conceptual framework. In Sect. 5 we propose a clasification of C^3. Related work is discussed in Sect. 6. Finally, we draw our conclusions in Sect. 7.

2. DEFINITIONS

We relate to the concepts of *Problem Space* and *Solution Space* from Jackson [10], the phenomena of the Problem World (PW, also called Problem Space or PS) and the phenomena at the interface PW- Machine (also called Solution Space), to examine the way in which aspects originate. We provide a semi-formal example that explains the way in which C^3 originates from the design decisions and modularisation concepts available to transform requirements into models and finally code. On the one hand, we have the requirements in the form of use cases that are modelled with class diagrams. The example is a web garment store we introduce in Sect. 3; on the other, we elaborate the specifications from the use cases and classes.

We consider that the selected framework, namely problem frames, allows discerning requirements (in the problem space) and specifications (in the solution space). This allows us to relate aspects through a tracing relation in terms of requirements that are transformed into specifications at later development phases, this distinction is valuable to clarify our definition and classification of C^3 and

*This work was carried out during the tenure of an ERCIM "Alain Bensoussan" Fellowship Programme

aspects.

In the following, we recall the definitions of requirements from Sommerville [16] and problem space from Jackson [10].

Definition 1 ***(Requirements)*** *"User requirements are statements, in a natural language plus diagrams, of what services the system is expected to provide and the constraints under which it must operate."*

Definition 2 ***(Problem space)*** *The problem space is defined as the set of requirements together with the explicit, or implicit, definition of the environment in which how the system to be should operate.*

Definition 3 ***(Solution space)*** *The solution space is the set of artifacts that belongs to a running implementation of a software system and without which the system would not be able to operate or would not have been produced (requirements are explicitly not in this set).*

In the web store example, the solution space consists of structural diagrams such as the class diagram shown in Fig. 1. as well as specifications.

Definition 4 ***(Transformation from problem space to solution space)*** *Given the notion of refinement as the transformation of an abstract, i.e., high level specification into one or more concrete or low level executable software artifacts, we defined the transformation from problem space to solution space as the process of transforming a given requirement into one or more executable artefacts by means of refinement. This mapping is performed explicitly, or implicitly, with the help of a given conceptual model that supports the translation of requirements to software models and finally to code.*

Moreover, solution and problem spaces are two concepts related by the transformation process from one to the other. We may formally relate the mapping either to a transformation function from problem to solution space; Sect. 3 provides an example of such transformation.

Some requirements are implemented in several modularisation units. This suggests a way to define crosscutting as a functionality or constraint characterizing more than one (logical at design, physical at code) module from one abstraction level to the next. This can be better understood by considering an abstraction level as a refinement step in a series of steps from requirements to code, being the previous step the most abstract in relation to the next one, such a relation is clearly transitive. We build a more abstract definition of C^3 than the ones explored so far by considering behaviour (functionality) as a property affecting several other entities, where a property is defined as a set of behaviours, so that an execution of a system Π satisfies a property P if and only if the behaviour (a sequence of states and agents) that represents the execution is an element of P see Abadi and Lamport [1]. This property can be implemented in several modules. In other words, *more than one* modularisation entity is present. The modularisation entity depends on the chosen architectural paradigm, e.g., object orientation, components, agents, or other.

The above reasoning leads to our definitions of crosscutting concern and aspect.

Definition 5 ***(Crosscutting concern)*** *Given a problem space a C^3 is a requirement that under some translation from the problem space to the solution space is expressed in more than one modularisation unit in a lower level of abstraction.*

We differentiate C^3 from *aspect* as follows. Crosscutting is due to the translation from the problem space to the solution space. This is due to the fact that no modularisation abstraction is perfect, yet aspects provide an additional modularisation unit to implement such crosscutting in models or code. Emphasizing this idea, aspects exist at the software architecture, design and implementation stages whereas C^3 can be seen as a superordinate concept, that is, a more generic one. This makes the aspects a subset of crosscutting concerns.

Definition 6 ***(Aspect)*** *Given a solution space an aspect is a specification (in some language, for instance the formal language Alloy [9] containing predicates, functions and facts as design entities) that represents a module with behaviour crosscutting other modularisation units (e.g., class, component, function); with respect to the underlying architectural framework.*

In the next section we introduce the language we use to express specifications, later used in our transformation case study.

3. AN OBJECT MODEL: WEB STORE DESIGN

The on-line garment store we present here was first developed by BOC Consulting GmbH. A modified version of this web store by Ioshpe [8] was used for an application of a security analysis. We later referred to Ioshpe's version in [6] and introduced it as a case study for a formal model of composition of security aspects. The web store model. In [7], we complete and modify the model by adding the system requirements in the form of use cases.

Obtaining the Structural Model (Class Diagram) from the Use Cases. we obtain the system's structure and functionality from the use cases shown in the Appendix. Our method is straightforward. We take the actors and entities as candidate classes. Based on the actions, from each use case, we draw the corresponding relations and decide which candidate classes appear as classes in the diagram. These are basically design decisions. It is actually not the aim of this work to propose a software development methodology, we rather rest on the methodology of object orientation, and common practice.

We consider each a partial diagram as part of the whole and iterate once a subsequent use case refers to an already considered actor or entity. If necessary, the diagram is modified.

Obtaining Classes from Use Cases. We associate actors and entities (nouns) to preliminary classes. Afterwards, we revise the resulting diagram against the relations among classes. Thereafter, we look for class attributes in the form of nouns and qualifiers. The process iterates when a common instance in a subsequent use case contradicts, or enhances, the existing partial diagram.

The Customer buys (relation) from the Office (web store) by selecting articles and placing an order. Given its kind of autonomous

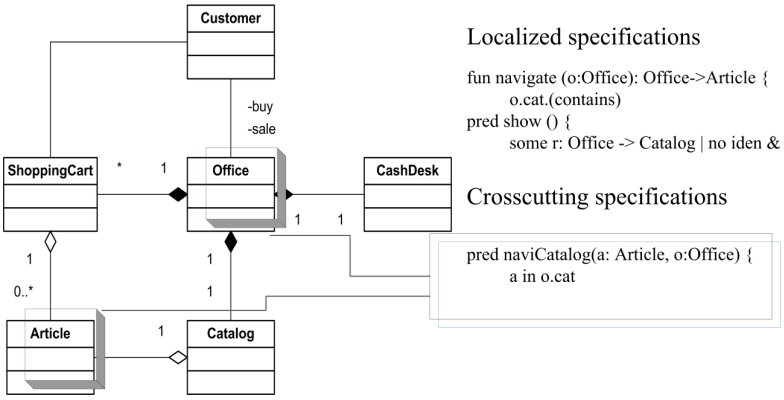


Figure 1: Classes and specifications from requirements

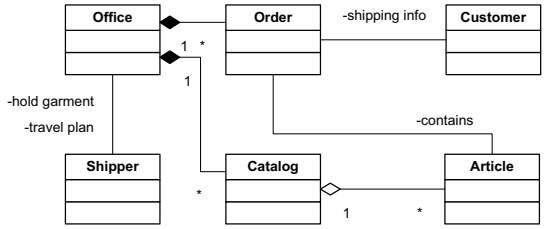


Figure 2: Classes from Use Case Process Order

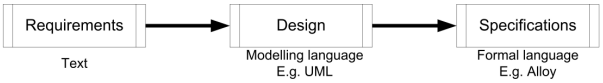


Figure 3: Transformation from requirements into specifications

Listing 1: Predicate navigate catalog

```
pred naviCatalog(a: Article, o:Office)  ↩   1
    {
        a in o.cat
}                                           3
```

existence, the Order is made into a class. The Office manages the product Catalogue and shipment process (class Shipper). This way we obtain the classes for the use case.

We associate actors and entities (nouns) to preliminary classes. Afterwards, we revise the resulting diagram against the relations described in the use case among classes. Thereafter, we look for class attributes in the form of nouns and qualifiers. The process iterates when a common instance in a subsequent use case contradicts, or enhances, the existing partial diagram.

For instance, the actor customer is mentioned a number of times in the use cases. The customer buys (relation) from the office (web store) by selecting articles and placing them in a shopping cart. Given its kind of autonomous existence, the shopping cart is made into a class. Also in this use case, garment (article) selection is completed by choosing: size and colour. These we incorporate as properties of article so we make them to attributes of this class. The resulting diagram is shown in the left hand side of Fig. 1.

We progress similarly along the following use cases and reiterate whenever a class, relation, or attribute, is contradicted or enhanced. This process terminates when the last use case is translated into a partial class diagram and then we elaborate upon the complete structural diagram by assembling the partial diagrams together.

The behaviour is obtained in another round of iterations. These are written in the form of specifications together with the structure, particularly as signatures, predicates and facts in Alloy.

We complete the structural model by setting the partial diagrams together. If necessary, we would iterate.

Specifications of our web store model are built upon a relatively easy to understand formal language supported by predicate, relational and "navigational" calculus. The objective is to provide a case study on which we may illustrate a premise. Namely that crosscutting concerns are requirements formulated in the problem space. Also, that after being translated into concepts of the solution space (e.g., modularisation units such as objects or logical components), these concerns can be traced to more than one such unit in the solution space.

In our web store case study, aspects are specifications (Alloy functions, facts, and predicates) that are related to a set of classes. As illustrated in Fig. 1, the predicate `naviCatalog` (see Listing 1) affecting classes `Office`, and `Article`. In contrast to function `navigate` in Listing 2 that operates over class `Office`. This translation from requirements into design entities and later to specifications is illustrated in Fig. 3.

In the following, we identify some types of C^3, characterise the sources of crosscutting, and present our classification of aspects.

Listing 2: Function browse catalog

```
fun navigate (o:Office): Office-> ↩
    Article {
        o.cat.(contains)
}
```

4. SOURCES OF CROSSCUTTING

4.1 Crosscutting Due to Inherent Limitations in a Decomposition Paradigm

As we already mentioned, crosscutting is due to the limitations of existing programming as well as design paradigms. Particularly the case of object orientation. The building block in object orientation is the class, which consists of data and methods. Some functionality mentioned in the requirements may be common to several classes. The methods implementing some requirements are therefore spread amongst several classes. For instance, the predicate specifying the navigation of the catalog in our web store case study. Compare this to a function for the customer to provide his name and address This function operates only on class *Customer*.

In other words, systemic concerns such as those that relate to a group of classes (such as security concerns like access control) are implemented into methods of several classes.

As another example think of this as a set W containing all the methods that realize access control. To further illustrate this, assume we have classes `client`, `server`, and `audit`, on which methods `client.authentication`, `server.authentication`, and `server.authorize` operate.

$$W = \{client.authentication, server.\text{-}authentication, \ldots, server.authorize\}$$

Where W relates to classes `Client`, `Server`, and `Audit`.

The reason for systemic concerns is that no programming paradigm is capable of allowing for a "concern-oriented" modularisation. This relates to the fact that the abstractions shaping a software system are constrained by the underlying modelling or implementation paradigm. For instance, the class/object concept.

Modularisation abstractions cause C^3. This is explored in Aksit [2] and more so in depth by Clarke in [3]. Particularly, Clarke demonstrates that the units of modularisation in object-orientation are structurally different from the units of modularisation in requirements specification. This result can be generalized to other modularisation paradigms. We express this notion in Definition 5.

4.2 Crosscutting from the Transformation of Non Functional Requirements into a Functional Proxy

Another source of crosscutting can be found in the translation of Non-Functional Requirements (NFR) into a functional implementation or a functional proxy of the corresponding Non-Functional (NF) specification. These are also considered in Definition 6. For instance, in [14] we present the translation of the security constraint *Keep transaction secure* in the early requirements stage into a number of security sub-constraints such as *keep transaction private, keep transaction available* and *keep integrity of the transaction* in the late requirements stage. These derived constraints represent the functional proxy of the more abstract security requirement.

5. A CLASSIFICATION OF CROSSCUTTING CONCERNS

We classify C^3 based on a classification of requirements since we define C^3 them as implementations of requirements.

Functional vs. non-functional C^3. Requirements are classified in *Functional (FR)* and *Non-functional (NFR)*. We understand FR as defining characteristics of the problem space and NFR as constraints in the solution space such as requirements on the system's performance. The behaviour of the SuD is expressed as FRs whereas restrictions to the possible solutions are defined in the form of NFRs or Design Requirements. Nevertheless, non-functional requirements are at some point in the development cycle translated into functional specifications. Meaning that they are translated into quantifications or behaviour. As an example, consider a NFR such as "Fault Tolerance". This requirement can, in subsequent phases of development, be translated into functional specifications that guarantee data persistence in view of a system failure. From this we obtain a first classification of C^3 as Functional and Non-Functional.

5.1 Non-functional Crosscutting Concerns or Aspects

We mentioned before that a NFR like *Fault tolerance* is subsequently translated to more concrete specifications that may finally affect a number of entities, think about a component or set of components guaranteeing data persistence in view of a system failure. More specifically, we might have a set of routines to ensure committing information (order number, article, price, and quantity) to the database once entered. Considered broadly, the implementation of such a requirement is dispersed through the three layers of the system yet predominantly in the database manager and in the application layer. In case we may draw a clear borderline of the entities affected, we identify non-functional aspects as systemic (NFS) or semi-localized (NFSL).

NFS are aspects identified by a boolean function "operates on" of the form: $OperateOn : Component \times Aspect \mapsto \{1, 0\}$, where component stands for modular entity. The function gives true (1) in case the component implements the aspect at least partly, and false (0) otherwise.

NFS. We define a NFS as an aspect that stems from a NFR and the function OperateOn gives true for more than half of the modules. In the case of the above example, we equate modules to layers. For example, "fault tolerance" implemented by a set of routines related to all three layers of the system and therefore represents a systemic non-functional aspect.

NFSL. We define a NFSL as an aspect that stems from a NFR and the function OperateOn is true for less than half of the modules. For instance, users of the Web Store access the system remotely, a requirement like keep information secret to third parties demands adding an encryption protocol between front end and application layer, yet we may consider it is set only at the communication links, not affecting other parts of the system.

As an example of C^3 from non-functional (or extra functional) requirements, consider a requisite establishing a given response time in use case "Payment process". This requirement may imply testing the system after constructing it to corroborate the system does not exceed a given response time from the moment the client provides his payment information and the confirmation from the office to the client the is received. This condition is related at least to classes: Customer, Office, Bank, and Account in our Webstore.

5.2 Functional Crosscutting Concerns or Aspects

These are aspects that arise from functional requirements, and are expressed in terms of behaviour rather than constraints. These constitute Functional Aspects (FA).

We explain these by the following example. The sub-requirement "keep transaction secure" is derived from the requirement as "Credit card information should be kept confidential, specially passwords". This requirement can be implemented by an encryption protocol. This aspect is in relationship to several other modularisation units.

Moreover, FA's can be implemented at the level of the methods in the classes, or at the level of behavioural specifications in components. For instance, black-box aspects are related to the public interfaces of components like functions, object methods, and communication channels. The behaviour affecting several modularisation units can be related to these via some wrapping over the communication channels or the external interfaces. Such as those that are capable of being implemented by frameworks as Composition Filters. Moreover, clear-box aspects relate to the internal structure of classes or components. In this case, the aspects can be seen as additional units composed inside the base ones. Jacobson introduced these aspects, years before the inception of the concept of AO, as "Existion and Extensions" in [11].

6. RELATED WORK

A review of other classifications reflects the predominant role of the programming level approach towards aspects. See for instance the work of Clifton [4]. The author proposes that aspects be categorized into two sorts: "spectators" and "assistants" with relation to the behaviour of the code that they advise. This classification is certainly of interest for programmers. Moreover, Rinard et al. [15] classify aspects based on the interaction between advice and method. Their classification helps programmers to understand the possible interactions of a given aspect and its design implications.

A categorisation of aspects at the programming level related to classes of temporal properties can be found in Katz [12]. The author defines the following classes of aspect: spectative, regulative and invasive. We share a common interest in providing aspects with semantics, yet it is at a different abstraction level. Marin et al. [13] propose a framework based on crosscutting concern sorts intended for aspect mining techniques. These sorts are defined as atomic descriptions of crosscutting functionality. Classified based on "intent" and its relationship to an aspect mechanism at the programming language level. We agree that the authors present a classification that supports aspect mining though very much influenced by the actual programming level constructs. This means that the elements they mine might not necessarily cover aspects in general though certainly those that they mention are *a priori* defined as C^3. In contrast, we define crosscutting and aspect and provide a classification that is independent of language mechanisms.

Similar work from van den Berg and Conejero [17] discussed crosscutting, though in terms of *tangling* and *scattering*, which are not yet not related to requirements traceability. Later on, van den Berg et al. [18] relate their definitions to requirements traceability, although they propose no classification and focus on identification of crosscutting in the early phases of software development. Our first exploration of the ideas introduced in this chapter can be found in [5]. In this work, we improved the definitions and enhanced previous work with concrete examples. Particularly at the hand of the web store system model that introduces a transformation example from requirements to specifications that was needed to illustrate our concepts. We also obtained a more clear classification of C^3 and aspects by categorizing them based on the type of requirement from which an aspect stems and the length to which the aspect relates to the system.

7. CONCLUSIONS

The discussion in this work provides an in depth investigation on the topic of C^3 and aspect orientation. Despite the popularity of AO and the number of tools, approaches and modelling techniques available to support one form or another of AO, regardless of some more or less precise definitions on *crosscuting* or *aspect*, some questions remained open. Hence, a more systematic study of the subject was needed. Particularly, from a perspective independent of the programming level. There are formalizations in the literature yet mostly, if not always, focused on the programming level. We pointed out that the subject itself cannot be defined simply by the programming constructs that embody what has been predominantly understood as aspect orientation, such as: *join points and advices*. These constructs are basically code transformation constructs that together with some sort of composition mechanism, that is, *weaving* allow enhancing a given base code.

Equally important, by exploring AO at the programming level we observed that the tools considered aspect oriented provide some sort of transformation mechanism that is applied on a given program using particular language constructs. This is also the case of aspect orientation at more abstract levels such as modelling. However, if we explore the subject in view of the question: how does aspect orientation make the world better? We face a difficult task answering, if we only defined the subject by the tools themselves. Therefore, we approached the topic from a top down perspective, bearing also in mind the intuitions taken by surveying the programming level techniques. We identified two main issues as being at the core of aspect orientation: *crosscutting* and *weaving*. We define crosscutting at the hand of the transformation from (informal) requirements to subsequent stages of development, namely their translation to a design or specification. We concluded that no modularisation entity may fully encapsulate the implementation of some requirements, and some features might be optional. Here appears the second " aspect" in AO, *weaving* as the composition mechanism that allows for defining a kind of abstraction layer over the base model or program and then inserting it at predefined points in the modules of the base model that should be enhanced. By describing crosscutting in view of the tracing or transformation from requirements to design or specifications, we provided a meaning to aspect orientation and the tools represented by it. We defined crosscutting as the cause and aspect as the solution to C^3. We further identified two main groups of C^3: *Nonfunctional* and *Functional*. Since aspects are the implementation of C^3 we focused onto the second group: functional aspects.

Finally, we conclude that the concepts of AO can be better explained at the software architecture level.

8. REFERENCES

[1] M. Abadi and L. Lamport. Composing specifications. *ACM Trans. Program. Lang. Syst.*, 15(1):73–132, 1993.

[2] M. Aksit. Composition and separation of concerns in the object-oriented model. *ACM Computing Surveys*, 28A(4), 1996.

[3] S. Clarke. *Composition of Object-Oriented Software Design Models*. PhD thesis, Dublin City University, 2001.

[4] C. Clifton. *A design discipline and language features for modular reasoning in aspect-oriented programs*. PhD thesis, Iowa State University, 2005. ISU TR 05-15.

[5] J. Fox. A taxonomy of aspects in terms of crosscutting concerns. In E. Brinksma, D. Harel, A. Mader, P. Stevens, and R. Wieringa, editors, *Methods for Modelling Software Systems (MMOSS)*, number 06351 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum für Informatik, 2007.

[6] J. Fox and J. Jürjens. A framework for analyzing composition of security aspects. In E. Brinksma, D. Harel, A. Mader, P. Stevens, and R. Wieringa, editors, *Methods for Modelling Software Systems (MMOSS)*, number 06351 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), 2007.

[7] J. Fox and A. G. Tornes. *Requirements Engineering with Aspects: Applying Aspect Orientation to Requirements in Software Development*. Lambert Academic Publishing, 2010.

[8] A. Ioshpe. Anwendung modellbasierter sicherheitsanalyse. Systementwicklungsprojekt, Technische Universität München, January 2004.

[9] D. Jackson. Alloy: A logical modelling language. In D. Bert, J. Bowen, S. King, and M. Waldén, editors, *ZB 2003: Formal Specification and Development in Z and B*, volume 2651 of *Lecture Notes in Computer Science*, pages 629–629. Springer Berlin / Heidelberg, 2003.

[10] M. Jackson. Problem frames and software engineering. *Information and Software Technology*, 47(14):903–912, Nov. 2005.

[11] I. Jacobson. Use cases and aspects -working seamlessly together. *Journal of Object Technology*, pages 7–28, 2003.

[12] S. Katz. Aspect categories and classes of temporal properties. In A. Rashid and M. Akşit, editors, *T. Aspect-Oriented Software Development I*, volume 3880 of *Lecture Notes in Computer Science*, pages 106–134. Springer, 2006.

[13] M. Marin, L. Moonen, and A. van Deursen. A common framework for aspect mining based on crosscutting concern sorts. Technical Report TUD-SERG-2006-009, Delft University of Technology, 2006.

[14] H. Mouratidis, J. Jürjens, and J. Fox. Towards a comprehensive framework for secure systems development. In E. Dubois and K. Pohl, editors, *Proceedings of the 18th International Conference, CAiSE 2006*, volume 4001, pages 48–62. Springer Berlin / Heidelberg, June 5-9 2006.

[15] M. Rinard, A. Salcianu, and S. Bugrara. A classification system and analysis for aspect-oriented programs. In *SIGSOFT '04/FSE-12: Proceedings of the 12th ACM SIGSOFT twelfth international symposium on Foundations of software engineering*, pages 147–158, New York, NY, USA, 2004. ACM Press.

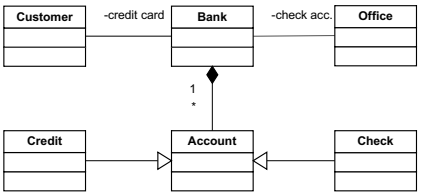


Figure 4: Classes from Use Case Payment Process

[16] I. Sommerville. *Software Engineering*. Addison-Wesley, 7th edition, 2004.

[17] K. van den Berg and J. M. Conejero. A conceptual formalization of crosscutting in AOSD. In *DSOA'2005 Iberian Workshop on Aspect Oriented Software Development*, Technical Report TR-24/05. University of Extremadura, 2005.

[18] K. van den Berg, J. M. Conejero, and J. Hernández. Analysis of crosscutting across software development phases based on traceability. In *Workshop in Aspect-Oriented Requirements Engineering and Architecture Design*, Shanghai, 2006.

APPENDIX

Use Case Payment Process

Description: The customer gives his credit card number and the office verifies his account through the bank. The bank debits the amount due and transfers the money to the web store's account.
Related Actors: Customer, Office, Bank
Precondition: Shopping cart completed. Preliminary invoice is ready. Delivery Address or pickup point is set.
Trigger: Closing the delivery menu.
Post-condition: Amount invoiced debited from the customer's account. This amount is transferred to the web store's account. Signal with "Okay" or "try again" sent to customer.
Standard Process:

1. The customer provides card number and related information in the web form.
2. The office sends these data to the bank, together with due amount.
3. The bank checks the card information.
   (a) If the information is correct, the amount due is debited from the customer's account. The web store becomes a confirmation and signals the customer "okay."
   (b) If the information is not correct, the bank sends a "try again" message. Up to three tries are allowed. Afterwards the process is aborted.
4. The office (web store) confirms the result of the process to the customer.

We identified in this use case the classes shown in Fig. 4. however after reading the corresponding use case in search of attributes it became manifest that account should be treated as an abstract class with two concrete classes: this was decided considering the difference between a checking account and a credit card account with respect to the attributes that each may contain. Therefore, class checking accounts have attributes regarding account numbers and balances, while credit card accounts have attributes regarding card numbers and credit limits.

Prediction and Validation of a Micro Wind Turbine for Rural Family Use

Narayanan Komerath
School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
Padma Komerath
SCV Inc, Johns Creek, GA 30022, USA

Abstract

This paper summarizes the development of a vertical axis micro wind turbine for families in remote areas. The design philosophy, requirements and constraints are summarized. Development steps are presented, including construction, static testing and measurement of performance. A numerical simulation captures the essential features of the design. Test results indicate moderate success with a 1meter diameter, 1 meter tall turbine using inexpensive and commonly available materials and construction techniques which allow use of a yawed biplane blade design. A 2m x 2m design scaled up from the 1meter machine requires stronger blade construction.

Introduction

This paper presents progress towards developing a micro renewable energy device optimized for use by families in areas remote from urban industrial centers. The design takes into account the constraints of access to repair facilities and parts, investment funding, family incomes, device attrition, and the paramount need for safety in environments where children and curious pets abound. Issues such as sustainability, use of commonly available off-the-shelf components, local buy-in, generation of local employment, and clean disposal of broken machines, must be considered in addition to thermodynamics and aerodynamics. This overriding philosophy dictates choices of materials, construction techniques and moving parts. Accordingly, the resulting device is less efficient in pure technical terminology than one that might be optimized for an industrial or military application. We will find that the Figure of Merit, defined as the ratio of what is achieved to the best that could theoretically be achieved, is much less than 1. This provides the technical challenge and motivation for research in this otherwise well-trodden field of endeavor.

The paper takes a somewhat convoluted path, for reasons tied to the above. We first explore the historical development of wind energy devices, and explain the motivation based on the difference between an optimization that might be done for the market in a developed nation, versus what might suit users in a place with very different realities. Next we discuss the design parameters, constraints, their rationale, and a brief history of the lessons leading to the present configuration.

Wind Energy Devices

The idea of extracting useful energy from the wind has been in vogue at least since the ancients put up a sail and set out across bodies of water [1,2]. There is mention of sails used in Egypt as early as 3200 BCE. Efforts to adapt this technology to stationary devices on land must have come in parallel or later [3-6]. One early implementation was a vertical-axle machine with rectangular sails driving a machine that ground grain, used in Sistan in the 8th century CE. There is evidence of wind turbines made of fabric and wood being used in China and Africa, with some of those technologies still in use in parts of Africa. These typically had wide-chord frames holding sails. Such devices were used for raising water from wells for irrigation, and perhaps also for grinding grain. The horizontal-axle wind turbine became a common landscape feature in Europe starting in the 12th century. Early versions were driven by aerodynamic drag force. More recent versions had airfoil blades that operated on lift, metal gears and machines powered by wooden-frame blades mounted on stone-built structures that doubled as the residence and factory of the mill operator. The Dutch windmill is universally recognized, and represents a highly successful element of the European rural economy of past centuries. In aerodynamic terms, this device is very inefficient. The rotor reaches barely above treetop level (there were not many trees around such windmills) and was at the bottom of the atmospheric boundary layer, thus getting only weak winds compared to today's tall towers. The stubby stone mill posed a large obstruction compared to today's slender towers. However, the devices worked. They delivered power directly to the point of use, and hence did not require conversion to electricity, or transmission through lines. The tower was integrated with the user's (very noisy) home. Most routine operations and maintenance were done by the residents, and related trades no doubt provided a stream of employment for locals.

There are two basic approaches to designing a wind turbine. One is to operate the device such that the blades generate aerodynamic drag, and are all pitched or twisted to a high angle so that a component of this drag force drives the device around the tip path circle. These devices are easy to build from wood or metal, and they operate over a wide range of wind speeds. With some modification, the blades would also generate some lift, but with a poor lift-to-drag ratio. This design was chosen, mounted on wood or metal towers, for the wind turbines that provided mechanical power to miners and farmers in the American West long before the railroad and the electric grid reached them. They are still used whenever one needs a quick and easy design and can build it out of metal sheets with minimal demands on skill. A less common variant of the drag-based design is the Savonius turbine, where the axis of rotation is perpendicular to the wind direction. At its most basic, this design consists of the two halves of a cylindrical barrel, cut along a diameter, and attached facing in opposite directions with some overlap at the middle, to

the shaft. It works on the principle that the drag on the side where the convex surface meets the wind, is about 1/2 to 1/3 of the drag on the side where the wind blows into the concave side. A cup anemometer is a form of Savonius turbine.

Most wind turbines today operate as lifting rotary wings, reaching lift-to-drag ratios above 60. This is much more efficient, and has led to the proliferation of ever-larger horizontal-axis wind turbines (HAWT, axis aligned with the wind direction) as primary wind-power extractors. The Darreus design of vertical axis wind turbines (VAWT) also uses a lift-based design. Thus it can achieve higher efficiency than the Savonius type, but at a considerable cost in complexity.

One complication is that a Darreus type machine cannot start itself as the wind starts from zero speed, and in fact operates most efficiently when the rotor tip speed is 3 to 5 times the wind speed. Thus large vertical axis wind turbines are typically started using an electric motor and taken to a relatively high speed. The power required for acceleration drops as the turbine picks up power from the wind, and the motor becomes a generator.

Design Parameters and Constraints

Our focus is on small machines suitable for operation by families in remote off-grid communities. When the machine becomes small, its aerodynamic efficiency is poor for various reasons. Low Reynolds number, low inertia, higher friction losses, and the low, fluctuating wind speeds available near the ground are obvious reasons. The cost of the non-power-extracting components becomes a larger part of the total, so that the cost per unit installed power becomes high. One unfortunate fact in dealing with commercial advertisements for micro renewable power devices is that actual performance rarely comes anywhere near the claims. Micro wind turbines, though relatively simple, pose installation issues and operational safety hazards [7] and may not survive weather conditions that are quite probable at least once every year. Many small commercial wind turbine models come with "rated power" values that exceed the physical limits of kinetic energy flux through their swept areas at the "rated wind speed". As may be expected, the reality of their operational performance is nowhere near these claims. The real mechanical figure of merit on most such systems, compared to the theoretical Betz efficiency of 59% of the kinetic energy flux for horizontal axis machines, is on the order of 10%. The electrical conversion efficiency is also low because of the wide range of torque and power encountered. It turns out upon careful investigation that the marketers of such devices, unlike those of large wind turbines, are not obliged to cite the rated power at the rated wind speed. The rated wind speed (typically around 12 m/s) is just a number suggested by the government. The rated power is what the manufacturers claim at the maximum wind speed (perhaps 40 m/s) that they claim the device will survive! Given that power is proportional to the cube of wind speed, one can easily calculate the consequences to optimistic customers who buy these systems. This is an increasingly strong cause of market resistance [8] to micro renewable power devices.

In reality, micro power devices need not compete with utility-scale devices on marginal cost per unit energy. Where the efficient power grid is not accessible, the real value [9] and the cost [10] of competing alternatives (mostly batteries) for the first several watts of installed power, or of the first several watt-hours of energy, are 3 to 4 orders of magnitude above the utility energy cost paid by grid-connected urban customers in developed nations [11]. Diesel generators are not much cheaper.

In this low-power regime of 10 watts to 1 kilowatt, we project that the vertical axis machine can be made competitive through research, and will offer real value to people in off-grid communities [12]. The blades of this machine encounter a large periodic fluctuation in aerodynamic load through each revolution, and in fact only a small portion of each blade's travel actually generates positive power. So the analysis and design of such machines are quite complex [13], even without the kinematic mechanisms used on commercial models to optimize blade loads through a cycle. However, these machines can be located close to the ground, and are hence more portable and accessible. They survive bad weather (they can be moved to shelter or dismantled) and high winds better. The centrifugal stresses are small compared to those on the blades of the horizontal axis machine, and the consequences of blade failure are far more benign. For these reasons, we focus on VAWTs.

In our laboratory, a 1m diameter x 1m high VAWT was originally conceived as a half-scale model, but was then seen to be useful in its own right. It is intended to be portable, and to allow a user to set it up easily outside his residence or on the roof. The power level is not anticipated to be above 100 watts mechanical. Accordingly the constraints are:

1. All rotary bearings are to be from bicycle components. The bearings are likely to be the highest cost items if custom-specified. However, bicycle components are familiar to residents, especially children, all over the world, and hence this provides a route for community "buy-in" of the maintenance of the device. An initial small-scale concept that constructed from a bicycle wheel is shown in Figure 1.
2. The blades must not be high-cost items. In fact we aim to make the blades easily disposable. In a practical environment, the blades are likely to be hit by objects such as farm tools, children's toys or balls, or falling branches. These things should not become catastrophically expensive terminators of the useful life of the device.
3. The blades should be bio-degradable in the long term and the products of their decomposition should be environmentally friendly. This rules out metal blades and sophisticated carbon fibers, which may not degrade. Many plastics are ruled out since they produce toxins if burned. To answer a reviewer's concern, "biodegradable" does not imply something that will dissolve in the first rain or become brittle in a year's worth of sunshine. Wooden parts are considered biodegradable, but have been used on ships and buildings for ages. On the other hand, many plastics are disastrous when dumped in backyards and landfills. Thus it is harmful to provide such "advanced" materials to areas and populations that lack the regular garbage removal, sorting/separation/recycling systems and tough waste disposal law enforcement of American suburbia.
4. Blade-making techniques must be compatible with generating semi-skilled jobs using local materials and labor. This helps "buy-in" and reduce cost of ownership.
5. The machine must "fail gracefully" and exhibit only benign modes of failure. Thus if a blade comes loose or breaks, it should not become a dangerous sharp-edged missile or flying sword. While studies on helicopter blade failures indicate that blade pieces quickly become unstable in flight and flutter down, there is no assurance that this will always occur in a failure situation. This makes it difficult to accept wooden blades without protective netting.

Having noted the above, we now go on to describe the research efforts. At this stage, blade materials and blade-making techniques are chosen with different objectives than those that will drive the final production versions. The first tentative experiments use the materials and designs that minimize cost and delay for small numbers of blades, and yet allow us to learn lessons that can be used in the design.

Early on, we recognized that the cheap-blade approach allowed us to attempt a biplane blade design. We conducted static wind tunnel tests to determine the lift and drag curve slopes of the blades, and how these were degraded when two blades were in proximity. The above constraints drove us to test flexible blades held stiff by tension. We believed that these would be amenable to production using textile and mat-weaving skills [14].

Early tests showed that incorporating a guide vane upstream of the machine would allow it to self-start, albeit slowly. A split 50mm diameter tube incorporated near the axis to serve as a Savonius starter has worked best to-date. Initial tests were conducted using a 1meter, 4-armed device with flexible double window-blind slat blades on each arm (Figure 2). These allowed generation of up to 260 rpm in the wind tunnel at 64 kilometers per hour wind speed. However, they degraded and tore in time, and this occurred rapidly when the machine was under load. Keeping both blades on each arm in tension proved to be difficult with the flexible blades, despite our developing special floating bracket attachments for the purpose. Going to a 3-armed version reduced the vibrations that were induced by resonant antisymmetric loading of symmetrically placed blades. For instance, consider what happens when identical blades pass through the 0 – 180 degree positions of the VAWT with respect to the wind direction. The blades at the front and back experience aerodynamic loads in the same direction. One adds to the centrifugal load whereas the other relieves the load, so that the net force on the machine is antisymmetric, setting the machine into a flapping mode of vibration. The flexible blade approach offered advantages for the 1m machine, but failed with a double-scale (2m x 2m) turbine (Figure 3) as discussed below.

A third-generation blade design used S2027 (low-speed, laminar-flow) airfoil section templates cut from plywood, with a PVC pipe acting as the spar and providing some torsional rigidity (Figure 4). PVC roof flashing sheets were bent and glued to the ribs, and stapled or glued together along the trailing edge. The blade roots terminated in PVC pipe fittings that are easily set at the designed blade pitch angle. While PVC sheets and pipe are not environmentally friendly, this technology can be used to demonstrate prototype machines, and get buy-in due to the easy availability and maintainability of the components. The blades of a VAWT generate driving torque and hence power, primarily when the blades are in two sectors of their orbits. To minimize the periodic fluctuation in power as the blades turned, the blades were yawed as seen in Figures 4 and 5.

Our other wind turbine prototype is a 2x version of the smaller turbine, reaching 2m diameter and 2m height. This still has a small enough footprint to be placed outside a rural or urban home, and the components are small enough to be transported, for instance in a Sport-Utility Vehicle or pickup truck, though probably not in a small car. In many parts of the world, pieces that large are routinely transported a few at a time, strapped alongside bicycles. This device is believed to be capable of 1 KW power. The initial design still used bicycle bearings, but this may be changed either to motorcycle bearings, or to use multiple bearings from bicycles. The efforts to operate it at low wind speeds in the low-speed diffuser of our wind tunnel (See Figure 3), failed. The Savonius tubes started the device and reached a slow speed, but the flexible blades went into uncontrolled flutter and generated net drag. The device actually worked better when the "lifting" blades were removed and the machine was operated as a Savonius turbine. The purpose of this device has been redefined. It is now viewed as something that should be designed to survive very high wind speeds, permitting it to be left, for instance along a sea wall or beach ridge along the Pacific Northwest coast [15] where strong winter winds are expected. Accordingly, the blades of this machine have been redesigned with steel tubes, foam core covered with fiberglass sheets, and metal ribs. Blade construction is shown in Figure 5. Wind tunnel testing is scheduled for mid-2011.

Results

A DC generator was initially coupled to the VAWT. However, this generator was optimized for 5000 rpm, and at the 100-250rpm speeds that were reached by the turbine, its conversion efficiency was poor. The generator allowed us to directly power a compact fluorescent bulb equivalent to a 40-watt incandescent, but this only drew 6 watts and hence does not constitute any great demonstration of success. Given the difficulties in the power conversion, we decided to focus on mechanical power extraction. Hub-based power generators for bicycles are still too expensive [16] for many applications of the 1m turbine.

A rope dynamometer was used to maintain a constant torque on the device, using a pulley and weights. Thus only the shaft rpm has to be monitored to obtain mechanical power readings. A laser non-contact rpm counter was tried for initial tests. Subsequently an optical shaft encoder connected to a USB power supply was installed at the top of the shaft. Because this was difficult to install and was failure-prone, it was replaced with a magnetic shaft encoder. This permits the detailed time-variation of rpm to be recorded on a laptop computer, suitable for remote transmission via the internet. A constant load of 2 lb (08.9N) is applied to the rope dynamometer. Of course a production version of the machine will not have such instrumentation.

In the spring semester of 2010 the VAWT went through a series of tests. Material static bending strength tests were used to develop the low-cost blades. Results are shown in Figure 6. Pipes made of CPVC, available easily for home use, provide a good alternative to ordinary PVC and to wood, as shown in Figure 6. In order to reduce the deflection of the blades in the VAWT, a gardening spike was placed in the center of the CPVC tube. After reassembly of the reinforced blades, the VAWT was tested over a range of wind-speeds ranging from 10 to 44 MPH. The increased bending stiffness of the blades improved performance. Video monitoring of the turbine operation helped analyze blade deformation through the power-generation area, located approximately 120° of azimuth counter-clockwise (i.e., along the direction of rotation) from the front of the VAWT. The blades bend inwards in this region, with the inner blades having higher deflection so that the spacing between the two blades increases with wind speed. The blades of the 1m VAWT survive wind speeds over 80 kmph, but over time, the surfaces deform and the blade aerodynamic performance degrades. Figure 7 shows the turbine in operation, with the deflection of the blades visible.

Horizontal Axis Wind Turbine Tests

A "400-watt" horizontal axis wind turbine (HAWT) of 54 inch diameter was acquired for reference, and tested at the same location as the 1-m VAWT was tested. With a 50-Ohm resistor

as load, the power curve of the HAWT is shown in Figure 8. The rotational speed of the HAWT was limited electronically. Clearly the power output under these conditions is very small. One reality of the HAWT is that at rated power, it must operate at an rpm of between 500 and 1000 rpm, posing the dangers cited before. Such an installation would only be suitable well above reach of children and pets, implying a tower or rooftop installation.

Numerical Simulation

Initial efforts at numerical simulation of turbine performance used a blade-element formulation to construct the change in the velocity diagram with azimuth. The simulation was conducted using MATLAB coding. This was used to confirm that we were calculating the expected variation of power extraction with tip speed ratio. The peak extraction was found to be at tip speed ratios close to 5, although values above 3 were good. Values above 2 were essential to really start seeing good aerodynamic lift.

This blade element simulation was carried forward to test finite aspect ratio blades, and to see the relative effects of profile drag and induced drag. The results showed that aspect ratio has a high effect, since induced drag can become a large component of total drag for this ideal calculation. Lift curve slopes and profile drag values for this calculation were derived from the static wind tunnel tests performed with the blades.

The effect of the drag generators (for starting) was also incorporated into this simulation. Net power generation was calculated by averaging the power around the tip circle. However, the simulation results remained well above the measured power results. Typical results are shown in Figure 9. The abscissa is the tip speed ratio, between the machine tip speed and the wind speed. The power is in watts, and is far above any power measured to-date. The results show that the power curve is sharply peaked as expected for vertical axis wind turbines, and the peak occurs between tip speed ratios of 3 and 5. The parameter of interest in the particular study from which Figure 9 is taken, is the blade thickness, keeping the thickness to chord ratio constant. This means that aspect ratio decreases as the thickness (or the chord as indicated in the figure) changes. The purpose of this study was to check whether there is an optimal thickness or aspect ratio, because a larger thickness means blades of greater bending stiffness. The results show that the chosen baseline is close to optimum, both for peak power and for the peak occurring at the lowest tip speed ratio which is still near 5.

The above simulation also shows that the expected mechanical power of the 1m vertical axis machine is well below 100 watts at tip speed ratios lower than 2. This corresponds to observed results. The simulation lacks the detailed drag and interaction modeling required to accurately capture the experimentally found power. Figure 10 shows mechanical power measurements made using the 1m VAWT, using the rope dynamometer. The power levels are only on the order of 20 watts or less at 12 m/s, as expected.

Conclusions

The philosophy behind the design of small, low-cost vertical axis wind turbines for family use is explained, and leads to a design where blades are inexpensive and the rotating parts come from bicycles. The measured power levels are low, but these are justified by the cost of alternatives at these low power levels. Specific conclusions:

1. The figure of merit of small vertical axis wind turbines is low.
2. A double-bladed turbine design is effective where blade cost is low.
3. Flexible-blade designs are fairly effective for the 1m turbine in the 100 watt regime, but are inadequate for a 2m version of the design at any power level, due to flutter issues.
4. Blade element theoretical predictions of the power show that high rotational speeds are required to obtain good figure of merit where the turbine radius is small.
5. A simple self-starting vertical axis turbine design has been demonstrated.

Acknowledgments

The author is grateful for the efforts of several students at the School of Aerospace Engineering, Georgia Institute of Technology, and in colleges around the world for the efforts that have helped develop these designs through the years. This work is also partially supported through a grant from NASA under the EXTROVERT cross-disciplinary innovation initiative .

References

[1] Casson, L., *"Ships and Seamanship in the Ancient World"*. The Johns Hopkins University Press. pp. 57–58, 1995

[2] Heyerdahl, T., Sandweiss, D.H., Narvaez, A., *Pyramids of Tucume*. Thames and Hudson Inc., New York, 1995

[3] Dietrich Lohrmann, "Von der östlichen zur westlichen Windmühle", Archiv für Kulturgeschichte,77,1,1995, pp.1-30.

[4] Sathyajith, Mathew, *Wind Energy: Fundamentals, Resource Analysis and Economics*. Springer Berlin, 2006. pp. 1–9

[5] A.G. Drachmann, *Heron's Windmill.* Centaurus,7,1961.

[6] Lucas, Adam, *Wind, Water, Work: Ancient and Medieval Milling Technology*. Brill Publishers, 2006 p. 105.

[7] Anon, "AWEA Small Wind Turbine Global Market Study, Year Ending 2008", American Wind Energy Assoc., 2009.

[8]Mach, W., "Wind Turbines - Do you need one at home?" Telegraph.co.uk, May 3, 2007.

[9]Gadgil, A.J., Greene, D.M., Rosenfeld, A., "Energy-Efficient Drinking Water Disinfection for Greenhouse Gas Mitigation". Proceedings of ACEEE 1998, August 23-28, 1998.

[10]Adee,S., "Empire Off The Grid". *IEEE Spectrum*, Aug. 09.

[11]Komerath, N., Venkat,. V., Halka, M., "Micro Renewable Energy Systems". *Proc. Atlanta Conference on Science and Innovation Policy*, Atlanta, GA, October 2009.

[12]Komerath, N.M., and Komerath, P.P., "Are Distributed Energy Systems Optimal in India?" *Proc. WAVES Conference*, Houston, TX, July 2006

[13]Kirke, B.K., *"Evaluation of Self-Starting Vertical Axis Wind Turbines for Standalone Applications"*. Doctoral Thesis, School of Engineering, Griffith University, April1998.

[14]Komerath, N., "A Campus-Wide Course on Micro Renewable Energy Systems." *Proc. ASEE National Conference*, College Park, TX, June 2009

[15]Komerath, N., "Micro Renewable Energy Systems As A Vehicle For International Awareness". *Proceedings of the ASEE National Conference*, June 2010.

[16]Heine, J., Oehler, A., "Testing the Efficiency of Generator Hubs". *Vintage Bicycle Quarterly* Vol. 3, No. 4, 2005. Reprinted at www.vintagebicyclepress.com. Seen 11/18/2010.

ADVANCES FOR ENVIRONMENTAL PROTECTION: BIOSURFACTANTS PRODUCED BY *Lactobacillus pentosus* FROM TRIMMING VINE SHOOTS AS ALTERNATIVE TO CHEMICAL SURFACTANTS

Ana Belén MOLDES
Chemical Engineering Department, University of Vigo
Vigo, 36310 / Campus As Lagoas, Spain

Xanel VECINO
Chemical Engineering Department, University of Vigo
Vigo, 36310 / Campus As Lagoas, Spain

Rosa DEVESA-REY
Chemical Engineering Department, University of Vigo
Vigo, 36310 / Campus As Lagoas, Spain

and

José Manuel CRUZ
Chemical Engineering Department, University of Vigo
Vigo, 36310 / Campus As Lagoas, Spain

ABSTRACT

Large-scale production of surfactants by fermentation of cellulosic material, like trimming vines shoots, would reduce greenhouse gas emissions in comparison with the greenhouse emissions produced during the synthesis of surfactants from petrochemicals. Moreover, Lactic acid bacteria, considered GRASS microorganisms, can be employed to obtain biosurfactants that are known to interact favorably with human skin, hair, and eyes, and that are readily biodegradable. Surfactin (commercial biosurfactant produced by *Bacillus subtilis*), is able to reduce the surface tension of water to 27 mN/m although its use is not very extended because it is not water-soluble. In comparison with surfactin, biosurfactants from *Lactobacillus pentosus*, obtained by fermentation of trimming vines shoots, only can reduce the surface tension of water from 72 to 55 mN/m, but it has a higher capability than surfactin and Tween 20 to stabilized gasoil/water emulsions. Besides, it was found that biosurfactants from *L. pentosus* have similar adsorption properties on sediments than tween 20 or surfactin.

Keywords: Trimming vines shoots*, L. pentosus,* biosurfactants, surfactin, sodium dodecyl sulfate, Tween 20

1. INTRODUCTION

Surfactants create emulsions by enabling the suspension of hydrophobic compounds (oil, hydrocarbons) in water. They act as dispersants or flocculants, enabling the suspension of solids, such as paint pigments, in liquid. These useful properties have fueled the growth of the surfactant market, mainly in the bioremediation of contaminated sites by hydrophobic compounds. Surfactant market has annual global sales of $23.9 billion [1] whereas the annual global production of surfactants is about 13 million metric tons [2, 3].

Surfactin is one of the most powerful known biosurfactants. It is able to reduce the surface tension of water from 72 to 27 mN/m at a concentration of 20 μM; however surfactin has limited utility for most consumer products applications because it is not water soluble and it has at elevated market price, 50 mg of surfactin from Sigma cost about 600 $.

On the other hand, trimmings of vine shoots from agriculture industry, related with wine elaboration, usually are field burnt; releasing cancerous compounds like polycyclic aromatic hydrocarbon as well as greenhouse gases. Therefore, the utilization of trimming vine shoots as carbon source to produce biosurfactants can decrease the environmental impact produced by this kind of residue, when it is field burnt.

Reznik et al. [4] proposed the utilization of cellulosic material, soy hulls, as carbon source to produce biosurfactants. These authors employed a genetically engineered gram-positive soil bacterium derived from *Bacillus subtilis* to produce the biosurfactant. Other authors have found that *Lactobacillus pentosus* can produce biosurfactants from trimming vine shoots [5, 6]. For instance, Bustos et al. [5] carried out continuous production of lactic acid and biosurfactants from hemicellulosic sugars of trimming vine shoots. However there are no much comparative studies of this biosurfactant from *L. pentosus* with other biosurfactants.

In this work *L. pentosus* is proposed to produce biosurfactants from trimming vines shoots, and a comparative study between biosurfactants from *L. pentosus* and commercial surfactants is done.

2. MATERIALS AND METHODS

Hydrolysis of trimming vines shoot

Ground samples of trimming vine shoots were hydrolysed under selected conditions (3 % H_2SO_4, 15 min, 130 ºC, liquid/solid ratio 8:1 g/g) and neutralized with $CaCO_3$ to a final pH of 6.5. The $CaSO_4$ precipitated, was separated from the supernatant by filtration.

Microorganism

L. pentosus CECT-4023 T (ATCC-8041) was obtained from the Spanish Collection of Type Cultures (Valencia, Spain). The strain was grown on plates using the complete media MRS agar.

Fermentation of hemicellulosic sugars from trimming vines shoots by *L. pentosus*

The clarified hydrolysates were supplemented with nutrients (10 g/L of yeast extract and 10 g/L of corn steep liquid), sterilized and used directly as fermentation media. The chemostat fermentation was carried out in a 2 L Applikon fermentor at 200 rpm with 1.6 working volume at 31 ºC and pH controlled to 5.85 during 48 hours.

Extraction of biosurfactants

Cells were recovered from the fermentation medium by centrifugation, washed twice in demineralized water, and resuspended in 50 mL of phosphate-buffer saline (PBS: 10mM KH_2PO_4/K_2HPO_4 and 150 mM NaCL with pH adjusted to 7.0). The bacteria were left at room temperature up to 2 hours with gentle stirring for biosurfactant release. Biosurfactants were obtained in the PBS and bacteria were removed by centrifugation. The remaining supernatant liquid was tested for surface activity.

Surface activity determination

The surface activity of biosurfactants produced by the bacterial strains was determined by measuring the surface tension of the samples with the ring method. The surface tension of PBS extract containing the biosurfactants from *L. pentosus* was measured using a KRUSS K6 Tensiometer equipped with a 1.9 cm Du Noüy platinum ring. To increase the accuracy an average of triplicates was used for this study.

CMC of bisourfactant from *L. pentosus*

The biosurfactant concentration was determined using a calibration curve mg/L of biosurfactants (surface Tension (mN/m)-72.6)/(-8.64). The calibration curve was calculated for a commercial biosurfactant produced by Bacilli (surfactin) using different concentrations of biosurfactant solutions, below the critical micelle concentration with known surface tension. In this biosurfactant concentration range the decrease of surface tension is linear and it is possible to establish a relationship between the biosurfactant concentration and the surface tension [7].

Surfactants

The surfactants employed in this work consisted on Tween 20; sodium dodecyl sulphate (SDS) supplied by PANREAC, surfactin supplied by SIGMA and biosurfactants from *L. pentosus* obtained in this work.

Surfactant sorption onto sediments

10 g of sediments (containing < 1% texture 63 μm; pH=6.1; 3.03 % of organic matter and a C/N ratio 12) were added to different solutions of surfactans (Tween 20, SDS, surfactine and biosurfactant from *L. pentosus*) at a liquid/solid ratio of 1:10 and the surface tension of the water was measured employing a KRUSS K6 Tensiometer equipped with a 1.9 cm Du Noüy platinum ring. Surface Tension was measured up to 200 min. To increase the accuracy an average of triplicates was used for this study. The concentration of surfactants employed in the sorption experiments was two times over their CMC.

Emulsification studies

Emulsification was performed according to the methodology proposed by Das, et al. [8]. Gasoil was vortexed during 2 minutes with an equal volume (2 mL) of aqueous phase containing surfactants and biosurfactants. After vortexing, tubes were left to stand for 1 h, and after that time (considered the initial time: 0 h) relative emulsion volume (EV, %) and emulsion stability (ES,%) were measured at 24 h intervals up to 72 h using the equations 1 and 2 proposed by Das et al. [8].

$$\text{EV}, \% = \frac{\text{Emulsion height, mm} \times \text{Cross section area, mm}^2}{\text{Total liquid volume, mm}^3} \qquad Eq.\ (1)$$

$$\% \text{ ES} = \frac{\% \text{ EV, at time h}}{\text{EV, at 0 h}} * 100 \qquad \text{Eq. (2)}$$

3. RESULTS AND DISCUSSION

From the adsorption experiments on sediments in can be observed that biosurfactants from *L. pentosus,* are adsorbed very fast; with similar behaviour to that observed for surfactin. Tween 20, a non-ionic surfactant; also was adsorbed by the sediments but the adsorption process was slower than that observed for biosurfactants from *L. pentosus* or surfactin, whereas sediments almost did not adsorb SDS, probably because SDS is an anionic surfactant. Consequently it can be speculated that biosurfactants from *L. pentosus* has similar adsorption properties to non-ionic surfactants like tween 20 or surfactin (the biosurfactant produced by *Bacillus subtilis*). **Figure 1** shows the evaluation of surface tension of water in presence of different surfactants and sediments. The increase of the surface tension of water is related with the adsorption of surfactants on sediments.

It can be speculated that the minor adsorption of SDS on sediments can be attributed to repulsion forces between the negatively charged SDS and the negatively charged sediments.

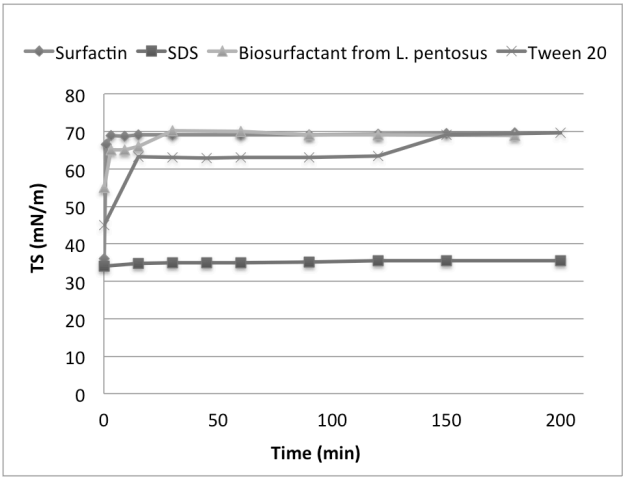


Figure 1. Evolution of surface tension of water in presence of sediments and surfactants two times over their CMC (critical micellar concentration).

On the other hand, **Figure 2** shows the capability of biosurfactants from *L. pentosus* to stabilize gasoil/water emulsions in comparison with SDS, tween 20 and surfactin. It can be observed that biosurfactants from *L. pentosus* gave better percentages of emulsions (EV=16 %) in comparison with surfactin (EV=0%) or Tween 20 (EV=9 % at 3.5 h and EV=7.7 % at 48 h); although SDS showed the best emulsify capability (EV=50%). Biosurfactants from *L. pentosus* and SDS gave emulsions very stable (ES=100%) whereas Tween 20 gave stability values (ES) = 85 % after 48 h.

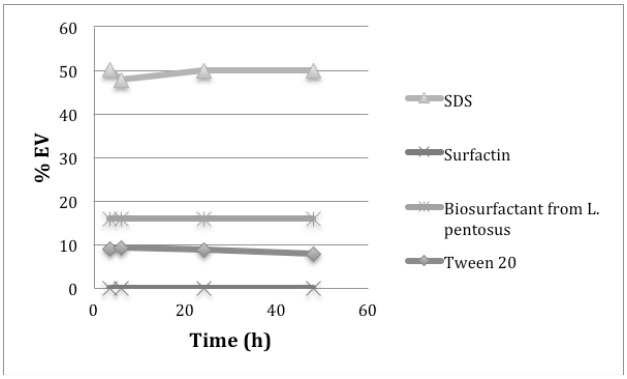


Figure 2. Percentage of gasoil/water emulsion formed in presence of different surfactants.

Related with the petrochemical consumption for the production of surfactants, Patel et al. [9] found that 0.57 tons of petrochemical intermediates were consumed per each ton of surfactant produced. Assuming that these ratios apply to global surfactant manufacturing today, surfactant production consumes about 7.4 Mt of petrochemical intermediates annually [4]. Life cycle analysis has estimated that each ton of petrochemical intermediate used for surfactant production generates 4,270 kg of emitted CO_2 [9]. Thus, the use of petrochemicals to produce surfactants generates annual emission of 31.6 billion kg of CO_2. According to the U.S. Environmental Protection Agency, combustion of one gallon of gasoline produces 8.8 kg of CO_2. Thus, Reznik et al., [4] estimated that annual worldwide use of petrochemicals for surfactant production emits CO_2 equivalent to the combustion of 3.6 billion gallons of gasoline. Consequently the commercialization of biosurfactants produced by fermentation of trimming vine shoots by *L. pentosus* could suppose an important advance for the environmental protection.

4. CONCLUSIONS

It is possible produce biosurfactants using *L. pentosus* from hemicellulosic sugars of trimming vine shoots, and this biosurfactants are adsorbed on sediments similarly to surfactin and tween 20. On the other hand biosurfactants from *L. pentosus* has better properties to stabilize gasoil/water emulsions than Tween 20 or surfactin.

ACKNOWLEDGMENT

We are grateful to the following institutions for the financial support of this work: Spanish Ministry of Science and Innovation (project CTM 2008-01608/TECNO) and to Xunta de Galicia Regional Government (Ángeles Alvariño Program) for financial support.

5. REFERENCES

[1] Market Report, **World Surfactant Market: markets, products, applications, innovations, chances & risks, competition, prospects to 2015**, Germany: Active Market Intelligence, 2008.

[2] D. Rust, **Surfactants: a market opportunity study update**, Midland, MI: OmniTech International Ltd., 2008.

[3] M.I. Levison, **Surfactant production: present realities and future perspectives**. In: U. Zoller (ed) **Handbook of detergents part F: production**, Boca Raton, FL: Vol. 142, pp. 1–38, 2009.

[4] G. Reznik, P. Vishwanath, M.A. Pynn, J. Sitnik, T. Jun Wu, Y. Jiang, B. Keenan, A. Castle, R. Haslell, T. Smith, P. Somasundaran and K. Jarrell. "Use of sustainable chemistry to produce and acyl amino acid surfactant", **Applied Microbiology Biotechnology,** Vol. 86, 2010, pp. 1387-1397.

[5] G. Bustos, N. de la Torre, A.B. Moldes, J.M. Cruz and J.M. Domínguez, "Revalorization of hemicellulosic trimming vine shoots hydrolyzates through continuous production of lactic acid and biosurfactants by *L.pentosus",* **Journal of Food Engineering**, *Vol.* 78, No. 2, 2007, pp. 405-412.

[6] M. Portilla, A. Torrado-Agrasar, J. Carballo, J.M. Domínguez and A.B. Moldes, "Development of a factorial design to study the effect of the major hemicellulosic sugars on the production of surface-active compounds by *L. pentosus",* **Journal of Agricultural and Food Chemistry**, Vol. 57, No. 19, 2009, pp. 9057-9062.

[7] S. Kim, E. Lim, S. Lee, J. Lee and T. Lee, "Purification and characterization of biosurfactants from *Nocardia sp* L-417", **Biotechnology and Applied Biochemistry***, Vol.* 31, 2000, pp. 249-253.

[8] M. Das, S.K. Das and R.K. Mukherjee, "Surface active properties of the culture filtrates of a Micrococcus species grown on n-alkenes and sugars", **Bioresource Technology**, Vol. 63, 1998, pp. 231–235.

[9] M.K. Patel, A. Theiss and E. Worrell, "Surfactant production and use in Germany: resource requirements and CO2 emissions", **Resources Conservation and Recycling**, Vol. 25, 1999, pp. 61–78.

Project Evaluation Using a Backpropagation Deep Belief Network

Alick NGUVULU
Graduate School of Information Science and Technology, Hokkaido University
Kita 14 Nishi 9, Kita-Ku, Sapporo 060-0814, Japan

Shoso YAMATO
Graduate School of Systems and Information Engineering, Tsukuba University
Tennodai 1-1-1, Tsukuba 305-8573, Japan

and

Toshihisa HONMA
Graduate School of Information Science and Technology, Hokkaido University
Kita 14 Nishi 9, Kita-Ku, Sapporo 060-0814, Japan

ABSTRACT

Recently a project evaluation approach called Project Assessment Indicator (PAI) Model has been developed based on expert knowledge. The PAI Model has been succesfully applied to evaluate overall monthly project performance based on 15 project elements derived from the project management (PM) processes and results groups. However, the evaluation process seems to be too subjective due to overdependency on PM experts to manually assign model weights. This study proposes using a Backpropagation Deep Belief Network (BP-DBN) to address this limitation. To this end, we use 7 real IT projects as our study objects. We completely eliminate the manual assignment of model weights in our BP-DBN approach which in effect increases the objectivity of the project evaluation process. The evaluation performance on the test data shows that our BP-DBN could compute monthly project performance based on the 15 project elements to within a relative error $|\varepsilon| \leq 3.3\%$ of the PAI Model value.

Keywords: Deep Belief Network, Machine Learning, Project Evaluation and Project Management.

1. INTRODUCTION

Project evaluation is a systematic analysis of information about project processes and results to determine the performance of a project. The evaluation outcome partially forms the basis on which Project Management Office (PMOs) make decisions to adjust the project processes for improved project performance and success. Thus it is important that PMOs use project evaluation approachs that are as rational and objective as possible.

Recently, a project evaluation approach called Project Assessment Indicator (PAI) Model [1] has been devised based on expert knowledge. The PAI Model takes 15 project elements as input variables and outputs a project performance indicator called PAI. The project elements are derived from the Project Management (PM) knowledge areas of scope, cost, schedule, quality, risk, procurement, communication and monitoring [2].

The merit of the PAI Model is its computational simplicity and ability to comprehensively measure the overall performance of any project based on the 15 project elements. However, evaluating projects using the PAI Model seems to be a too subjective process. This is in consequence of allocating model weights and category scores by PM experts based on the empirical rule. Category scores here refer to the maximum points each of the 15 project elements can score. We observe that a more rational approach for allocating weights and category scores is essential for increased objectivity of the project evaluation process and the project performance indicator itself.

Project evaluation is essentially a pattern recognition problem to which machine learning techniques like traditional neural networks (NNs) have been applied. For example in [3][4], NNs have been successful applied to predict project performance while in [6][5], NNs have been applied to model project risk and talent management.

The main focus of this study is to demonstrate that a Backpropagation Deep Belief Network [7] (BP-DBN) could provide a rational approach for: (i) automatically assigning weights to the project elements, (ii) computing category scores using the backpropagation weights, and (iii) objectively measuring project performance. A BP-DBN is simply a generatively pretrained DBN Model trained with teacher data using backpropagation to perform discriminative modeling. Although not yet applied in PM, DBNs and their variants have been used successfully in 3-D object recognition [8], road detection from aerial images [9], facial expression generation [10], reduction of the dimensionality of data [11] and image processing [12].

The rest of the paper is organized as follows. Section 2 gives an overview of some classical project evaluation methods followed by details of the PAI Model. Section 3 describes the basic principles of DBNs. Section 4 presents the data followed by the experimental setup in Section 5. Results are discussed in Section 6. Concluding remarks are given in Section 7.

2. PROJECT EVALUATION METHODS

We first briefly describe three widely used project evaluation methods: Critical Path Method (CPM) [13], Program Evaluation and Review Technique (PERT) [14] and Earned Value Management System (EVMS) [15] and then give details of the PAI Model later.

Critical Path Method
CPM is one of the useful scheduling tools applied in planning and controlling many interrelated activities in a large and complex project. The method consists first in the identification of the so-called critical paths, critical activities and critical events in the network, which is the project model. The next step is the computation of values useful for PMOs and Project Managers. The computed values contain such information as the slack of events and activities, the earliest and the latest times of the start and the finish of particular activities. CPM mainly focuses on project schedule control.

Program Evaluation and Review Technique
PERT is another network-based PM tool for planning, scheduling and controlling many interrelated activities in a large and complex project. PERT uses the same approach and network representation as the CPM to capture the precedence or parallel relationships among the activities in the project and to compute values that are useful to the PMOs and Project Managers. CPM and PERT are usually used in a combination often referred to as CPM/PERT. Again the focus of CPM/PERT is mainly scope and schedule control of the project.

Earned Value Management System
An EVMS integrates a project's work scope, cost, and schedule into a single performance measurement baseline (PMB) and reliably tracks the planned value of work, the earned value of actual work done, and the actual cost of work done. It compares actual performance measures with the PMB and provides the means for identifying and reviewing changes to the PMB. It further provides trend analysis and evaluation of estimated cost at completion and a sound basis for corrective actions, and management re-planning.

As can be seen, these classical methods mainly focus on the performance of project scope, schedule and cost with no direct reference to the other PM knowledge areas.

PAI Model
The PAI Model is a quantitative project evaluation approach recently devised by Yamato et al [1] based on expert knowledge. The model takes 15 project elements, shown in Table 1, as input variables and outputs an indicator of overall project performance called PAI. As earlier stated in the introduction, the 15 project elements are derivatives of the PM knowledge areas of scope, cost, schedule, quality, risk, procurement, communication and monitoring. For convenience's sake, we divide the elements into two main categories: PM Processes and PM Results (See Table 1). The PM Processes Category elements countercheck the quality of management mechanisms and tools adopted in managing a project. The PM Results Category elements countercheck the quality of the results of adopting management mechanisms and tools.

Table 1 - The PAI Model input variables divided into the PM Processes and Results Groups.

Project Element	PM Processes Group	PM Results Group
Financial	✓	✓
Customer	✓	✓
Review	✓	✓
Progress	✓	✓
Performance	✓	✓
Quality	✓	✓
Organization	✓	✓
Risk	✓	

Suppose that P is the total score of the PM Processes elements p_i and R is the total score of the PM Results elements r_i. We compute the overall project performance indicator PAI using the PAI Model as:

$$\mathrm{PAI} = \mathrm{P} + \mathrm{R} \tag{1}$$

where:

$$\mathrm{P} = \sum_{i \in M} w_i p_i;\ \mathrm{R} = \sum_{j \in N} v_j r_j \tag{2}$$

w_i is the weight assigned to the variable p_i, M is the number of variables in the PM Processes group, v_j is the weight assigned to the variable r_j and N is the number of variables in the PM Results group. Experts assign weights empirically and intuitively based on their PM skills and experience. The correlations between the project elements is not take into account when assigning weights w_i and v_j and category scores p_i and r_j. We observe that the project evaluation process using the PAI Model may be too subjective due to the human influence arising from the differences in the PM skills and experience of experts assigning the weights.

3. DEEP BELIEF NETWORKS

Deep Belief Networks (DBNs) are generative probabilistic models recently introduced within the emerging field of deep machine learning. Evolved from general belief networks [13], DBNs are known to have better knowledge representational capabilities than traditional NNs [14].

As illustrated in Fig. 1, the main building block of a DBN is a model called Restricted Boltzman Machine (RBM). An RBM consists of a layer of stochastic binary visible units $\mathbf{v} \in \{0, 1\}$ which interact with an upper layer of stochastic binary hidden units $\mathbf{h} \in \{0, 1\}$ via weight W. There are no connections between units of the same layer.

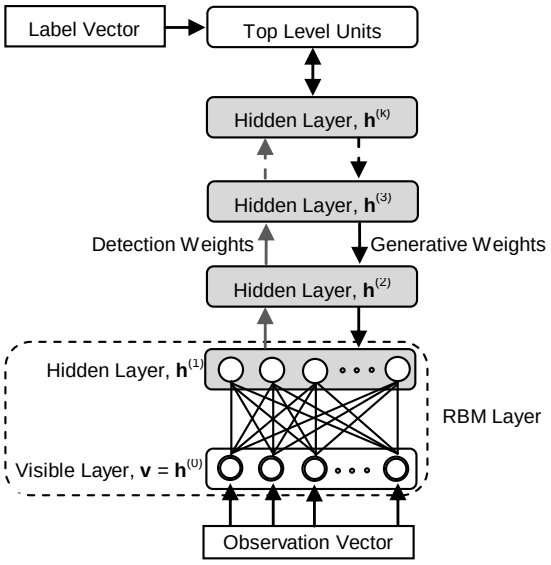


Figure 1: General framework of a Deep Belief Network.

The attractiveness of a DBN lies in its use of RBMs as model building blocks by which joint conditional probability distributions are determined over the observed variables $v_i \in \mathbf{v}$ and hidden variables $h_j^{(k)} \in \mathbf{h}^{(k)}$ where $k = 1,\ldots, \ell$. hidden layers. Denoting $\mathbf{v} = \mathbf{h}^{(0)}$, the joint distribution over the visible variables and hidden variables is given as:

$$p\left(\mathbf{h}^{(0)}, \mathbf{h}^{(1)}, \mathbf{h}^{(2)},..., \mathbf{h}^{(\ell)}\right) = p\left(\mathbf{h}^{(0)}|\mathbf{h}^{(1)}\right) p\left(\mathbf{h}^{(1)}|\mathbf{h}^{(2)}\right),..., p\left(\mathbf{h}^{(l-2)}|\mathbf{h}^{(l-1)}\right) p\left(\mathbf{h}^{(l-1)}|\mathbf{h}^{(l)}\right) \quad (3)$$

where all the conditional layers $p(\mathbf{h}^{(l-1)}|\mathbf{h}^{(l)})$ are factorized conditional distributions such that

$$p\left(\mathbf{h}^{(l-1)}|\mathbf{h}^{(l)}\right) = p\left(h_i^{(l-1)}=1|\mathbf{h}^{(l)}, \mathbf{W}^{(l)}, \mathbf{b}^{(l-1)}\right) \quad (4)$$

$\mathbf{b}^{(l-1)}$ is the bias vector of layer l-1 and $\mathbf{W}^{(l)}$ is the weight matrix representing the symmetric interaction terms between units in layer l-1 and units in layer l. Notice that the RHS of Eq.(5) is given as follows:

$$p\left(h_i^{(l-1)} = 1|\mathbf{h}^{(l)}, \mathbf{W}^{(l)}, \mathbf{b}^{(l-1)}\right) = \sigma\left(b_i^{(l-1)} + \sum_j w_{ij}^{(l)} h_j^{(l)}\right) \quad (5)$$

where:

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (6)$$

A DBN learns to represent knowledge using a greedy layer-wise unsupervised learning algorithm introduced by Hinton et al. [7]. Consider an observation vector $\mathbf{v}$ (hereafter denoted as $\mathbf{h}^{(0)}$ for the sake of consistency) presented to the visible units from where the input signals are transmitted to the hidden units. Going in a top-down direction, the hidden units $h_j^{(1)}$ attempt to reconstruct the original input vector by stochastically finding the visible unit inputs $h_i^{(0)}$. The new visible unit activations are forwarded to the hidden units where another attempt to reconstruct the original inputs is made. This back and forth Gibbs sampling process is repeated until one step reconstruction hidden unit activations $h_j^{(1)}$ are attained. During this learning process, also called contrastive divergence learning [15], the weights are repeatedly updated based on the difference between two pairwise correlations as follows:

$$\Delta w_{ij} \propto \langle h_i^{(0)} h_j^{(1)} \rangle_{data} - \langle \hat{h}_i^{(0)} h_j^{(1)} \rangle_{reconstruction} \quad (7)$$

where Δw_{ij} is the weight update, the correlation $\langle h_i^{(0)} h_j^{(1)} \rangle_{data}$ is the frequency measure with which the visible unit i and hidden unit j are both on when the state of the hidden units $\mathbf{h}^{(1)}$ is determined from the original inputs $\mathbf{h}^{(0)}$ using Eq.(9).

$$p\left(h_j^{(1)} = 1|\mathbf{h}^{(0)}, \mathbf{W}^{(1)}, \mathbf{b}^{(1)}\right) = \sigma\left(b_j^{(1)} + \sum_i w_{ij}^{(1)} h_i^{(0)}\right) \quad (8)$$

Similarly, the correlation $\langle \hat{h}_i^{(0)} h_j^{(1)} \rangle_{reconstruction}$ is the frequency measure with which the visible unit i and hidden unit j are both on when the state of the hidden units $\mathbf{h}^{(1)}$ is determined from the *reconstructions* $\hat{\mathbf{h}}^{(0)}$ of the original inputs $\mathbf{h}^{(0)}$ according to Eq.(10).

$$p\left(h_j^{(1)} = 1|\hat{\mathbf{h}}^{(0)}, \mathbf{W}^{(1)}, \mathbf{b}^{(1)}\right) = \sigma\left(b_j^{(1)} + \sum_i w_{ij}^{(1)} \hat{h}_i^{(0)}\right) \quad (9)$$

The *reconstructions* are computed as follows:

$$p\left(\hat{h}_i^{(0)} = 1|\mathbf{h}^{(1)}, \mathbf{W}^{(1)}, \mathbf{b}^{(0)}\right) = \sigma\left(b_i^{(0)} + \sum_j w_{ij}^{(1)} h_j^{(1)}\right) \quad (10)$$

This process of learning hidden variables could be repeated for any number of hidden units in as many layers as we desire. Once the generative pretraining is done, the DBN can then be fine-tuned through supervised learning to perform discriminative modeling. In supervised learning, label data are presented to the unit(s) in the top level layer of the pretrained DBN. The DBN is then trained using the well-known backpropagation algortihm and its convergence monitored through the minimization of either the mean squared error or the cross-entropy error functions. Notice that new weights are established between the top level layer and the last hidden layer while the pretraining weights are only slightly adjusted.

4. DATA

The experiment data are obtained from 7 real IT projects carried out from April 2006 to February 2007 (Table 2). Each project has all the 15 project elements evaluated in the months shown as well as the overall monthly project performance indicator (PAI) computed using the PAI Model. All but Projects B1106 and B1202 have 12 months durations. Since 12 months is the maximum project duration we set the 5 projects with 12 months duration i.e. B1101, B1102, B1104, B1203 and B1205 as training data. This implies that our training dataset consists of 900 cases of monthly imput data (i.e. 15 monthly input variables for 12 months for 5 projects) and 60 cases of monthly target data (i.e. 1 monthly PAI for 12 months for 5 projects).

Projects B1106 with 10 duration (monthly input data = 150; monthly target data = 10) and B1202 with 11 months duration

(monthly input data = 165; monthly target data = 11) are set as test data..

Table 2: Months (shaded) in which the project shown in the leftmost column were evaluated.

Project	2006									2007		
	Month											
	A	M	J	J	A	S	O	N	D	J	F	M
B1101												
B1102												
B1104												
B1106												
B1202												
B1203												
B1205												

Legend

Shade	Meaning
	Projects with 12 Months duration
	Project with 11 Months duration
	Project with 10 Months duration

5. EXPERIMENTAL SETUP

Normalizing Data

We normalize the experimental data via Eq.(11) such that the data have zero mean and unit variance.

$$y_i' = \frac{y_i - \mu}{s} \tag{11}$$

where y_i' is the normalized value of y_i, μ is the mean of the data and s is the standard deviation.

Pretraining the DBN

We use the greedy layer-wise contrastive divergence algorithm to train our DBN. The model parameters pertinent ot training the DBN are the number of hidden layers and units, weight initialization, weight-decay, the learning rate and momentum, and batch size [16].

Number of hidden units: Our BP-DBN consists of 15 units in the visible layer, 500 units in the first hidden layer, 15 units in the second hidden layer, 500 units in the third hidden layer and 15 units in the fourth hidden layer. The top-level layer consists of 1 unit corresponding to the output variable which is monthly project performance in this study. The 15 visible units correspond to the 15 input variables (the monthly project elements). Figure 4.1 illustrates the DBN Model we use in this study.

Initial values of the weights and biases: The RBM's are trained sequentially one at a time in a bottom-up direction. As shown in Fig. 4.1, RBM 1 is trained first and its output becomes the input to RBM 2. This is repeated for RBM 3 and RBM 4 until we get a complete pretrained DBN. Throughout all these steps, we initialize the weights to a normal distribution of $\mu = 0$ and $\sigma = 0.02$ while the hidden biases are set to 0.

Weight-Decay: Weight-decay is an extra term that added to the normal gradient to reduce overfitting to the training data and to improve generalization. In this study, we use a weight decay of 0.0001 throughtout our experiments.

Learning rate and momentum: We train each RBM in 220 epochs with a fixed learning rate of 0.01 and a momentum of 0.9 throughtout the experiments. Momentum is a simple method for increasing the speed of learning when the objective function contains long, narrow and fairly straight ravines with a gentle but consistent gradient along the floor of the ravine and much steeper gradients up the sides of the ravine.

Size of each mini-batch: For a more efficient update of the model weights, it is necessary to divide our dataset into small mini-batches of 75 cases which are used to train the RBMs through stochastic gradient descent.

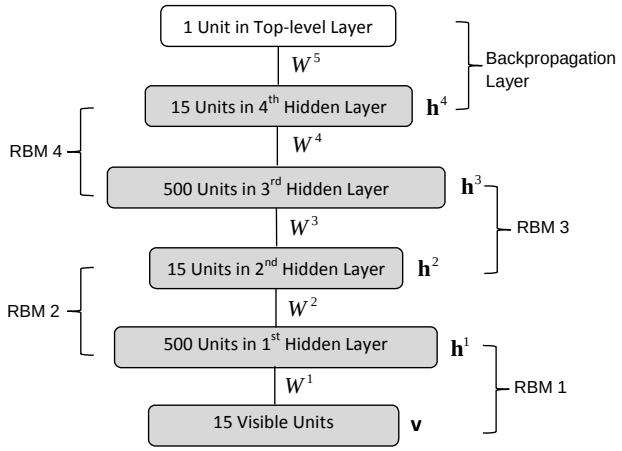


Figure 2: The 4-RBM layer BP-DBN Model consisting of 15 units in the visible layer **v**, 500 units in the first hidden layer $\mathbf{h^1}$, 15 units in the second hidden layer $\mathbf{h^2}$, 500 units in the third hidden layer $\mathbf{h^3}$, 15 units in the fourth hidden layer $\mathbf{h^4}$, and 1 unit in the top-level layer. The 15 visible units correspond to the 15 input variables. Notice that **v** interacts with $\mathbf{h^1}$ via W^1, $\mathbf{h^1}$ interacts with $\mathbf{h^2}$ via W^2, $\mathbf{h^2}$ interacts with $\mathbf{h^3}$ via W^3, $\mathbf{h^3}$ interacts with $\mathbf{h^4}$ via W^4 and $\mathbf{h^4}$ interacts with the top-level unit via W^5.

Training the DBN Using Backpropagation

In this phase, we introduce 1 unit in the top-level layer to train the DBN to perform our dscriminative task of modeling project performance using 15 input variables. The 1 unit corresponds to the output variable i.e. monthly PAI. We simultaneously present the input variables to the visible units and monhtly PAI to the top-level unit. At this stage, training reverts back to the usual backpropagation (BP) learning [17]. Hence our model name BP-DBN. Notice that the weights W^1, W^2, W^3, and W^4 from the pretraining phase are only slightly adjusted while a new vector of backpropagation weights W^5 linking $\mathbf{h^4}$ with the top-level unit are established. We use stochastic gradient descent with a mini-batch size of 75 as pretraining.

6. RESULTS AND DISCUSSION

We assess the ability of the DBN to model monthly project performance based on the 15 project elements using the relative error of estimation ε, defined as:

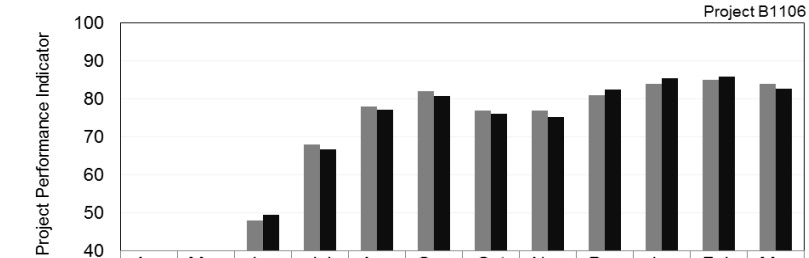
$$\varepsilon = \left(\frac{\left| T_i^{BP-DBN} - T_i^{PAI} \right|}{T_i^{PAI}} \right) \text{ x } 100\% \qquad (12)$$

where $T_i^{BP\text{-}DBN}$ and T_i^{PAI} are monthly project performance computed using the BP-DBN and PAI Model for month i respectively. Table 3 shows the maximum relative error of estimation for each project and the month of the maximum. Oberve that left side of Table 3 shows the training error $|\varepsilon_1|$ data (Projects B1101, B1102, B1104, B1203 and B1205) while the right side shows the test error $|\varepsilon_2|$ (Projects B1106, and B1202). Notice that $|\varepsilon_1| \leq |\varepsilon_2|$. We now present detailed comparisons of monthly project performance computed using the BP-DBN and the PAI model for Projects B1106 and B1202. On test data from

Table 3: Relative Error of Estimating Monthly Project Performance using our BP-DBN

Project	$\|\varepsilon_1\|$ **% (*Month*)**	**Project**	$\|\varepsilon_2\|$ **% (*Month*)**
B1101	1.95(*July*)	B1106	3.13(*June*)
B1102	1.95(*June*)	B1202	3.27(*November*)
B1104	1.88(*March*)		
B1203	1.90(*November*)		
B1205	2.03(*November*)		

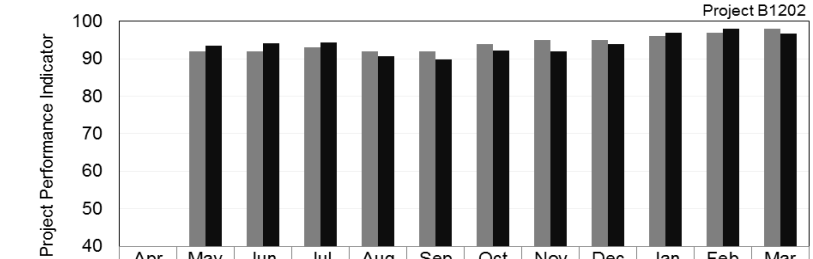
Project B1106, the BP-DBN computes overall monthly project performance within a relative error margin $|\varepsilon_2|^{B1106} \leq 3.13\%$ of the PAI Model computed value. On average the BP-DBN computes monthly project performance with a mean relative error $|\varepsilon_2|^{B1106} = 1.74\%$. Figure 4 shows the comparison of the BP-DBN computed monthly project performance indicators (Black bars) with the PAI Model computed values (Light Grey bars).



	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar
PAI Value			48	68	78	82	77	77	81	84	85	84
DBN Value			49.50	66.73	77.17	80.86	76.15	75.22	82.59	85.57	85.95	82.65

Figure 4: Monthly project performance indicators computed using the PAI Model (Light Grey bars) and the BP-DBN (Black bars).

On test data from Project B1202, the BP-DBN computes overall monthly project performance within a relative error margin $|\varepsilon_2|^{B1202} \leq 3.27\%$ of the PAI value. On average the BP-DBN performs with a mean relative error $|\varepsilon_2|^{B1202} = 1.73\%$. Figure 5 shows the comparison of the BP-DBN computed monthly project performance indicators (Black bars) with the PAI Model computed values (Light Grey bars). Comparing the relative errors $|\varepsilon_2|^{B1106}$ and $|\varepsilon_2|^{B1202}$ or the mean relative errors, we can see some consistence in the BP-DBN's computation of monthly project performance on the test data. This implies that when we use the BP-DBN trained on our current dataset to evaluate a project, we are likely to obtain monthly project performance values T_i, such that $T_i^{BP\text{-}DBN} = T_i^{PAI} \pm \varepsilon$ where $T_i^{BP\text{-}DBN}$ and T_i^{PAI} are as defined in Eq.(12) and $|\varepsilon| \leq 3.3\%$ is the relative error.



	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar
PAI Value		92	92	93	92	92	94	95	95	96	97	98
DBN Value		93.56	94.06	94.28	90.77	89.73	92.11	91.89	93.82	96.99	98.02	96.66

Figure 5: Monthly project performance indicators computed using the PAI Model (Light Grey bars) and the BP-DBN (Black bars).

7. CONCLUSION

To the best of our knowledge, the work reported in this paper is the first attempt at applying BP-DBNs to project performance evaluation. Our proposed approach shows that we could use BP-DBNs to compute reliable monthly project performance indicators based on the basic project elements data. The merit in this BP-DBN approach is that the model weights are automatically assigned during model training during with the possibility of self-adjustment as new data is presented to the operational model. This contrasts the subjective approach in the PAI Model where model weights are manually assigned and fixed by humans. We argue that not only can our BP-DBN evaluate projects comprehensively but also that it provides a more rational and objective evaluation approach.

ACKNOWLEDGEMENT

The authors would like to thank the Ministry of Education, Culture, Science and Technology (MEXT) of Japan and the Japan Society for the Promotion of Science (JSPS) Grant-In-Aid for Scientific Research (A)(20241042) for supporting this work.

REFERENCES

[1] S. Yamato and J. Nakamura, "Project Assessment Indicator and PMO Actions for Project Maturity Model," **3rd International Conference on Project Management National Conference**, CS10D, Sydney, 2006.

[2] S. Yamato, J. Nakamura and T. Honma, "Quantitative Project Assessment Using the AHP Absolute Measurement Method," **4th International Conference on Project Management**, ProMAC2008, Anchorage, USA, 2008.

[3] A. Nguvulu, S. Yamato and T. Honma, "Forecasting Project Performance Using a Neural Predictor Model", **IEEJ Transactions in Electronics, Information and Systems**, Vol. 131, No. 4, 2011, pp. 900-905.

[4] S. O. Cheung, P. S. P. Wong, A. S. Y. Fung and W. V. Coffey, "Predicting Project Performance Through Neural Networks", **International Journal of Project Management**, Vol. 24, 2002, pp.207-215.

[5] N. Goonawardene, S. Subashini, N. Boralessa, and L. Premaratne, "A Neural Network Based Model for Project Risk And Talent Management," In: L. Zhang, J. Kwok, and B. L. Lu (eds): **ISNN 2010, Part II, LNCS 6064**, 2010, pp. 532-539.

[6] S. A. Sarcia, G. Cantone, and V. R. Basili: "A Statistical Neural Network Framework for Risk Management Process – From The Proposal to its Preliminary Validation for Efficieny," In: **ICSOFT (SE)**, INSTICC Press, 2007, pp. 168-177.

[7] G. E. Hinton, S. Osindero, and Y. The, "A Fast Learning Algorithm for Deep Belief Nets", **Neural Computation**, Vol. 18, 2006, pp. 1527-1554.

[8] V. Nair and G. Hinton, "3-D Object Recognition With Deep Belief Networks", in **Advances in Neural Information Processing Systems**, Vol. 22, 2009, pp.1339-1347.

[9] J. M. Sussking, G. E. Hinton, J. R. Mollevan, and A. K. Anderson, "Generating Facial Expressions Using Deep Belief Networks", In V. Kordic (ed) **Affective Computing, Emotion Modeling, Synthesis and Recognition**, ARS Publishers.

[10] V. Mnih and G. E. Hinton, "Learning to Detect Roads in High Resolution Aerial Images", **European Conference on Computer vision**, 2010.

[11] G. E. Hinton and R. R. Salakhutidnov, "Reducing the Dimensionality of Data With Neural Networks", **Science 313**, 2006, pp. 504-507.

[12] K. Tanaka, "Statistical Learning Procedure in Loopy Belief Propagation for Probabilistic Image processing", In **International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce** (**CIMCA-IAWTIC'05)**, Vol. 2, 2005, pp.741-746

[13] R. M. Neal, "Connectionist Learning of Belief Networks", **Artificial Intelligence**, Vol.56, 1992, pp. 76-113.

[14] Arel, D. C. Rose, and T. P. Karnowski, "Deep Machine Learning – A New Frontier in Artificial Intelligence Research", **IEEE Computational Intelligence Magazine**, Vol.5, No.4, 2010, pp.13-18.

[15] G. E. Hinton, "Training Products of Experts by Minimizing Contrastive Divergence", **Neural Computation**, Vol. 14, 2002, pp. 1771-1800.

[16] G. E. Hinton, "A Practical Guide to Training Restricted Boltzmann Machines, Version 1", **Technical Report 2010-003,** Machine Learning Group, University of Toronto, 2010.

[17] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Representations by Back-propagating Errors", **Nature**, Vol. 323, 1986, pp.533-535

Mobile Robot Designed for Research and Education in Engineering

Salvador Ramírez, Noé Barrera, Juan Anzurez, Isidro I. Lázaro, Luis A. Torres
Facultad de Ingeniería Eléctrica,
Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México.
szavala@umich.mx, , a0001125x@correo.fie.umich.mx, j.anzurez@ieee.org, ilazaro@ieee-sco.org, latsalomao@ieee.org

ABSTRACT

Mobile robots have the potential to become the ideal tool to teach a broad range of engineering disciplines. Indeed, mobile robots are getting increasingly complex and accessible. They embed elements from diverse fields such as mechanics, digital electronics, automatic control, signal processing and programming. Moreover, they are attractive for students, as a motivation means to learn. This paper presents a mobile robot design, based on the PIC microcontroller family. It incorporates ultrasound sensors, a magnetic compass and encoders. The user can control and/or program the mobile robot to navigate while receiving visual information remotely. All this in an indoor environment. The control system's components and the main characteristics of the constructed robot are described. Thanks to its modular design, the mobile robot can be used in a wide range of activities like household, industrial and educational; as shown on the experimental results of this paper.

1. INTRODUCTION

Mobile robotics is still an open and challenging field with a big projection into the future and lots of possible indoor as well as outdoor applications that can help to improve industrial production and some aspects of the workers life quality [1]. Very interesting applications are currently being developed, and in a medium term, will be part of our daily life; see for example [2]. However, most of these applications are still on a research stage, so it is important that the knowledge generated by these research efforts is gradually incorporated to topics that students, mainly engineers, learn at university.

A mobile robotic platform is a robot for experimental and research in real environments. These vehicles have similar characteristics to robots for industrial or commercial applications. A robotic platform is intensively used in the development phase of research projects applied to robotic systems. For example, robotic platforms are used for testing and validating robot control architectures, or to examine autonomous or semi-autonomous navigation algorithms or as a way of testing different sensors, actuators and other electronic components. We can mention the following examples: robot Khepera [3], Nomad 200 [4], Magellan and B21R [5], widely used platforms for experimental research in the field of robotics.

The purchase of one of these robotic platforms is generally of high cost. Other problems are some difficulties associated with import procedures, proprietary software licenses, technical support, etc. For this reason, we chose to develop our own prototype.

This paper presents the design and construction of a mobile robotic platform of high performance, suitable for teaching and research on issues related to mobile robotics. It is of great use in robotic systems for static and dynamic environments, for example, development and validation of algorithms for robot navigation and control.

The paper is organized as follows: Section 2 presents the general structure of the developed mobile robot. Section 3 presents different operation modes. In section 4, the position control algorithm is described. The mobile robot testing is presented in section 5. Finally, conclusions are shown in section 6.

2. MOBILE ROBOT GENERAL STRUCTURE

This project was divided into two design stages: mechanical and electronic.

Mechanical design. The robot has a metal structure that provides support for all hardware. This metal structure has a rectangular shape of 9.64in (24.5 cm) of width by 10.24in (26 cm) of length with a rear structure that stands 3.15in (8 cm), as shown in Figure 1. The structure has four holes, one on each corner, where rods are placed in order to support the underlying control circuit platform.

The robot energy source consists of two batteries. One feeds the motors and the other the control circuits; the battery specifications are 12 Volts, 12 Amp-hour and 8 Amp-hour respectively.

The mobile robot was designed as a tricycle, as shown in Figure 2a. It is a front-traction robot, and each wheel is attached with a set of gears like the ones used for speed bikes; one on the wheel axis and another on the motor axis. Both held together with a chain link as shown in Figure 2b. On the back, we incorporate a free motion wheel that can rotate 360 degrees that functions only as support for the mobile robot.

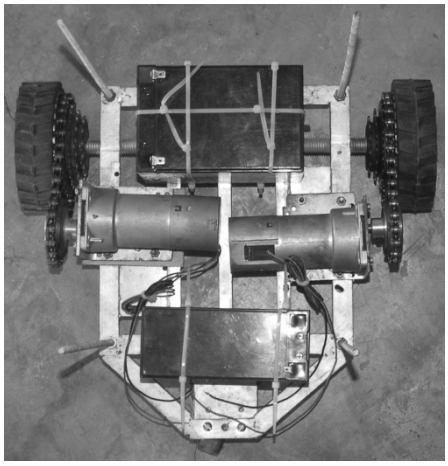
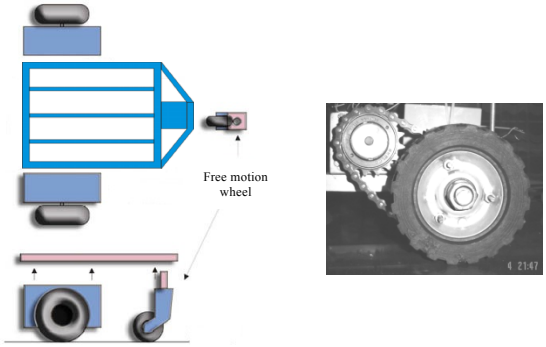


Figure 1. Metal structure



(a) (b)

Figure 2. (a) Wheel placing, (b) Motor and wheel coupling

Electronic design. The prototype was designed in a modular pattern, allowing different activities with slight modifications; Figure 3 shows a block diagram containing parts that correspond to the electronic design.

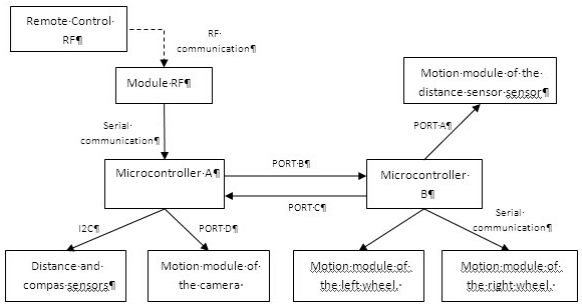


Figure 3. Electronic system's block diagram

The main subsystem is formed by a PIC16F877 microcontroller [6], [7], responsible for receiving information from the radio frequency module, the distance and compass, and to process and transmit movement signals from the camera and the mobile robot movements.

The communication protocols between these modules can be observed in figure 3. Parallel, serial and i2c communication protocols were used.

The secondary subsystem, formed by another PIC16F877 microcontroller, is responsible for driving the motors, moving the SRF08 ultrasonic sensor and is the link between the main subsystem and the control modules of each wheel.

The RF module receives operation commands to manipulate the movements of the robot and camera. It has a radio frequency receiver and transmitter (RWS434 and TWS315 respectively).

The system has two ultrasonic sensors (SRF08 and SRF02) and a compass sensor (CMPS03) [8], [9], [10]. The SRF08 sensor monitors the distance at the front of the robot (180 degrees of vision) for any object. The other sensor (SFR02) is used to detect steps on the ground, and is placed on an endless screw which rotates with the aid of a motor and moves the sensor to both sides of the robot sequentially. The compass sensor (CMPS03) provides orientation with respect to the Earth's magnetic north.

The camera motion module consists of: a small camera that transmits a video signal through RF and two coupled stepper motors for vertical as well as horizontal axis rotational movement in order to allow vision of the robot surroundings. This module is shown in Figure 4.

The remote RF control is shown on Figure 5. This remote control transmits command operations for robot and camera. The remote control includes a RF transmitter and receiver (RWS315 and TWS434). It also includes two manual controls and a personal computer and a microcontroller (PIC18F2431) which encode the sent data.

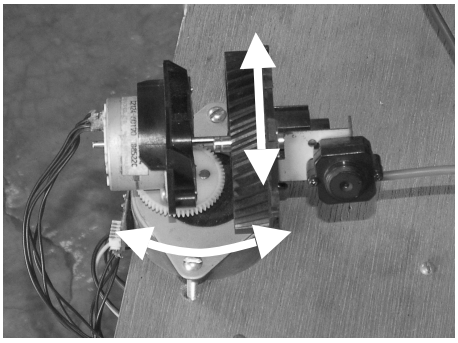


Figure 4. Camera placement

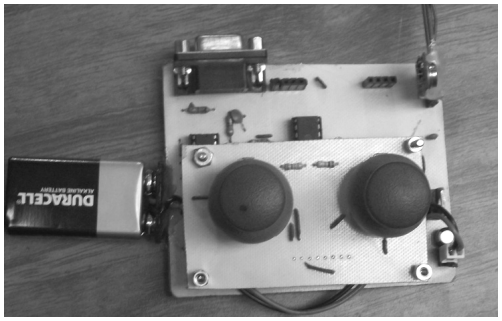


Figure 5. Wireless Remote Control

The mobile robot has a control module for each wheel. Each module includes a microcontroller (PIC18F4431), H-Bridge and optocoupler. For position and speed a digital PI algorithm is used.

Figure 6 shows a front panel picture of the mobile robot.

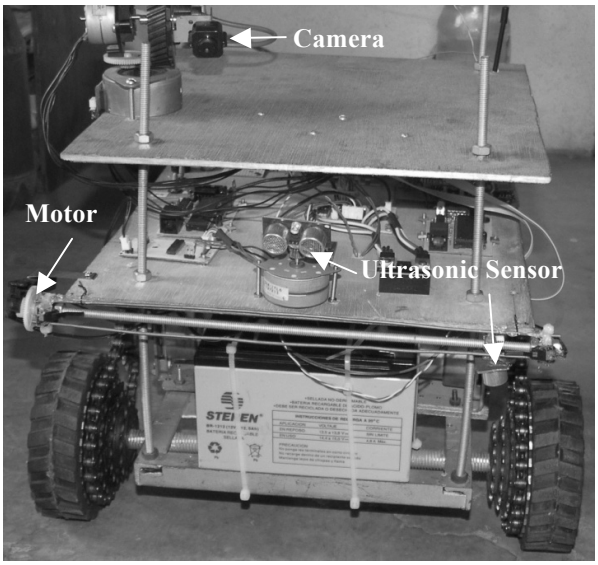


Figure 6. Mobile Robot front panel.

3. OPERATION MODES

The operation modes of the robot are shown in Table 1.

Table 1. Operation modes

Operation mode	Actions
Mode 1	Manual control of the robot.
Mode 2	Routine path unassisted.

In the operation mode 1, the user indicates the robot's motion using the left handle in the wireless remote control.

When any of the ultrasonic sensors detect any object or step that obstructs the robot's path, the secondary microcontroller, which controls motor motion, sends a stop command to both motors.

The stepper motors that hold the camera are controlled by the right handle.

In the operation mode 2, the user previously defines a desired motion sequence and subsequently transmits it to the robot.

Figure 7 shows an example of a motion sequence. Any sequence consists of a pair of commands whose format is: (*Orientation, Distance to go*). Orientation is a number from 0 to 255 that corresponds to 0 to 360 degrees. The distance ranges from 4-255 cm, as shown in Table II. Once the routine has started, the microcontroller's main task is to determine the type of obstacle in the way using the signal deciphered from the ultrasonic sensors. If the obstacle is small sized, the robot rounds it and continues with the defined sequence. In case the obstacle is big the robot aborts the sequence and returns to its initial position.

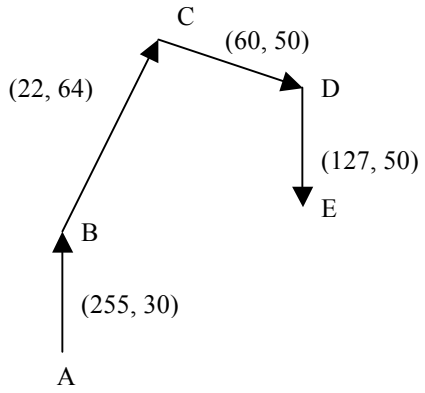


Figure 7. Path example

Table 2. Data from a sample trajectory corresponding to Figure 7

Path	Orientation degree. (byte)	Distance (cm)
AB	360^0 (255)	30
BC	31^0 (22)	64
CD	84.6^0 (60)	50
DE	179.1^0 (127)	50

4. POSITION'S CONTROL ALGORITHM

The objective of the control algorithm is to generate action signals such that the robot follows the desired directions. This objective can be achieved if each control module has a good position control of its motor's shaft. In order to control the position of each wheel, the speed profile shown in Figure 8 was selected [12].

In this profile, we can observe the following regions:

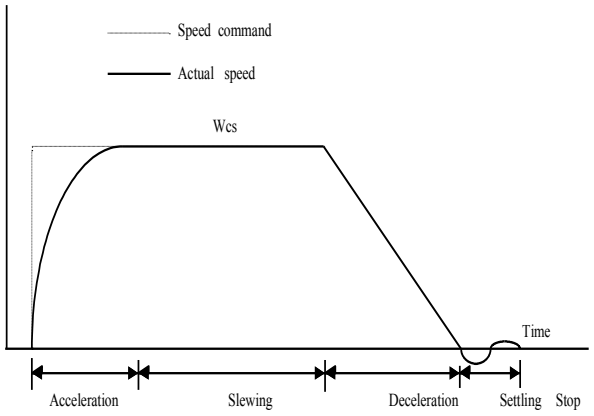


Figure 8. Speed profile

(a) Acceleration: The position error counter sends a command for maximum speed. The rotor is accelerated according to the dynamic equation of the motor and the speed controller is used.

(b) Slewing: When the rotor reaches the desired maximum speed, the speed controller maintains this speed constant.

(c) Deceleration: To decelerate the rotor with a constant torque, the command speed ω_{ref} is given according with (1).

$$\omega_{ref} = \sqrt{2\beta q} \tag{1}$$

where:

q = actual error position
β = deceleration rate

(d) Settling: Ideally, the motor must stop at the target position and must not travel any farther when control is governed exactly by (1). In practice, however, there is always a possibility that the rotor overruns beyond the target, because the speed command data is not given as an ideal linear function but discontinuously because of digital quantization. The system is normally designed so that the rotor eventually settles after several cycles of oscillation around the target.

(e) Stop: When the rotor comes to rest at the target, the controller sends the required torque to maintain the rotor at position. To get the speed profile shown in Figure 8, the controller generates speed commands ω_{ref} that depend on the position error in accordance to the following rules:

i) if $0 < q < q_f \Rightarrow \omega_{ref} = \sqrt{2\beta_{\max} q}$ (2)

ii) if $-q_f < q < 0 \Rightarrow \omega_{ref} = -\sqrt{2\beta_{\max}(-q)}$ (3)

iii) if $q_f < q < q_{\max} \Rightarrow \omega_{ref} = \omega_{\max}$ (4)

iv) if $-q_{\max} < q < -q_f \Rightarrow \omega_{ref} = -\omega_{\max}$ (5)

where:

q = actual error position
q_f = position error where the deceleration process starts
β_{max} = digital value that represents the maximum rate of deceleration
q_{max} = maximum position command

These rules are shown graphically in Figure 9.

5. RESULTS

Several tests were performed to verify the proper functioning of each module and its operation as a whole.

To test both modes of robot operation, the following test were made.

The first mode was checked by following the instructions from the remote control.

For the second mode, a defined path with and without obstacles was used (Figure 10). In this test satisfactory results were obtained with almost no tracking error in the trajectory.

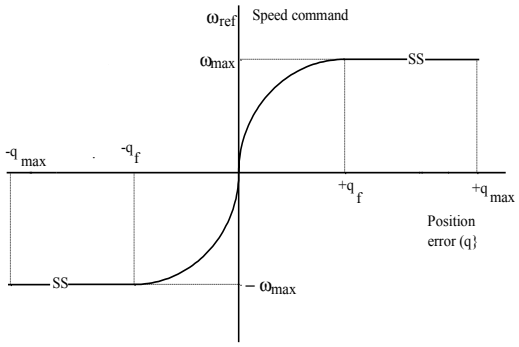


Figure 9. Speed commands in function of the position error.

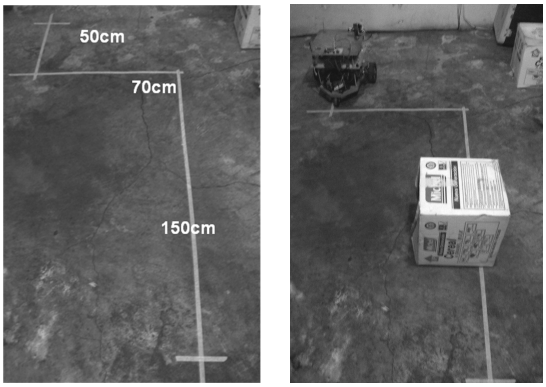


Figure 10. Trajectory with and without obstacles

Figure 11 shows the versatility of the robot in other activities. Other accessories can be included to clean using a defined path to sweep an area.

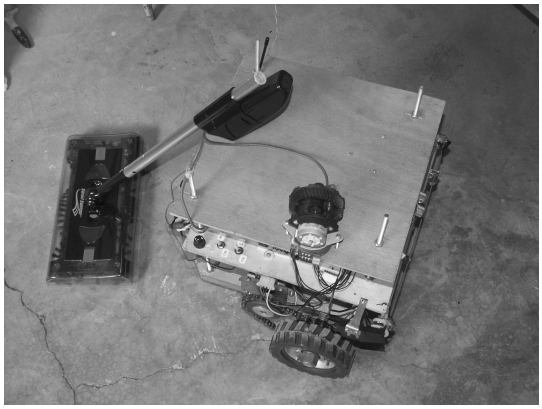


Figure 11. Cleaning robot

Finally, the robot was used as a watchdog. For this task, the first mode of operation was used, guided by remote control in an enclosed space.

6. CONCLUSIONS

In summary, this paper presented the design and implementation of a mobile robot. Several benefits can be mentioned, one of them, the opportunity for students to learn to make measurements and assemble a functional unit.

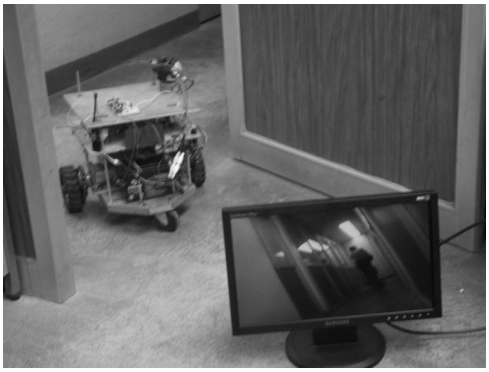


Figure 12. Watchdog robot

This paper illustrates two different ways of using the mobile robot as an incentive for teaching and learning.

The design of motion control systems for mobile robots is a challenging task. A reliable and powerful motion control system is the basic requirement for the robot to perform accurately. In this design, a digital motion control system was configured, designed, implemented, tested and evaluated. The adaptation of PID compensation and computational control gives the system maximum flexibility. Safety was emphasized as an important factor in the design.

The PID compensator is programmed inside the master controller of the mobile robot. It would be easy to reprogram this compensator using any other control technique for testing purposes. This shows how easily students could test their knowledge in the robotics control area.

The modular design of the implemented robot allows other modules like robotic arms, computers, vision transducers or other types of sensors to be added or taken off easily. Research would be a simple task because of the robot's versatility.

The mobile robot is under development and there are potential applications to be evaluated. Research for exploration and industrial inspection are possibilities yet to be explored.

7. REFERENCES

[1] A. Stentz, "Robotic Technologies for Outdoor Industrial Vehicles", in Proc. of SPIE AeroSense 2001.

[2] T.Pilarsky, M. Happold, H. Pangels, M. Ollis, and A. Stenz, "The Demeter system for automatic harvesting", Autonomous Robots 13, pp. 9-20, 2002.

[3] K-Team. Khepera user manual. Lausana: EPFL, 1998

[4] Nomadic technologies, Inc. http://www.robots.com.

[5] Real world interface. http://www.rwii.com.

[6] Angulo J., Microcontroladores PIC Diseño Practico de Aplicaciones, McGraw Hill 1997.

[7] PIC16F87xA microcontroller manual, http://www.microchip.com.

[8] Robodacta http://www.robodacta.com.mx/

[9] Santafe Y. E., Jugo D., Cote P., González M., "Sistema de detección de obstáculos por sonar ultrasónico para personas invidentes (Bastón ultrasónico)", Proc. II Congreso Colombiano de Ingeniería e Ingeniería Biomédica, Bogotá, October 2005.

[10] Herrera A., Jaramillo R., "Diseño y construcción de un dispositivo para la alerta de obstáculos (DAO)", Ingeniería Biomédica magazine, ISSN 1909-9762, number 1, may 2007, Pag. 23-27.

[11] Muhammad H. Rashid. Power Electronics Circuits, Devices and Applications. Prentice-Hall 1993.

[12] Takashi Kenjo, Power Electronics for the Microprocessors Age, Oxford Science Publications 1994, Pages. 311-325

BIOGRAPHIES:

Salvador Ramirez Zavala received his degree in Electrical Engineering at the Universidad Michoacana de San Nicolás de Hidalgo (UMSNH), México, and his Master´s degree in Electrical Engineering at the same institution in 1990 and 1998, respectively. He is now a full-time Professor/Researcher on the Faculty at the UMSNH. His areas of interest are Control Systems and Power Electronics.

Noe Barrera Gallegos was born in Morelia, Mich. He received his degree in Electrical Engineering at the Universidad Michoacana de San Nicolás de Hidalgo (UMSNH). He is now a full-time postgraduate student on the Faculty at the UMSNH. His areas of interest are Power Electronics and Robotics.

Juan Anzurez Marin received the B.E. degree in Electrical Engineering from the Universidad Michoacana de San Nicolas de Hidalgo (UMSNH), Mexico, in 1991; the M.Sc. degree in Electronic Engineering from the Instituto Tecnologico de Chihuahua, Mexico, in 1997 and Ph.D. degree in Electrical Engineering option Automatic Control from the Centro de Investigacion y Estudios Avanzados del IPN (CINVESTAV), Guadalajara Campus, Mexico, in 2007. He is an Associate Professor with Department of Electrical Engineering of UMSNH from 1987, Morelia, Mexico. His research interests include Instrumentation and Control Systems as well as the Fault Diagnosis Algorithms Design for the Nonlinear Systems and Energy Harvesting Applications. He is an IEEE member.

Isidro Ignacio Lázaro Castillo was born in Córdoba, Veracruz, Mexico. He received his degree in Electrical Engineering at the Universidad Michoacana de San Nicolás de Hidalgo (UMSNH), and his Master's degree in Electrical Engineering at the same institution in 1992 and 1999, respectively. He is now a full-time Professor/Researcher on the Faculty at the UMSNH. He has published various scientific and technical articles in national and international congresses and journals. He is the author of the book *Ingeniería de Sistemas de Control Continuo*. His areas of interest are Energy Quality, Power Electronics, Control Systems and Parameter Identification.

Luis Alberto Torres Salomao received his B.S. degree in Electronic Engineering from the Universidad Michoacana de San Nicolas de Hidalgo (UMSNH), Mexico. He is currently half-time professor at UMSNH and postgraduate student at Centro de Ingeniería y Desarrollo Industrial (CIDESI - CONACyT). His research interests are Control Systems, Fault Diagnosis, Microcontroller applications and Mecatronics. He is an active IEEE member.

Writing and Monitoring in International Standardization, Theoretical Choices and Methodological Tools

Brigitte Juanals
MoDyCo (UMR 7114) – Université Paris Ouest Nanterre La Défense – CNRS
Nanterre, 92001, France

and

Jean-Luc Minel
MoDyCo (UMR 7114) – Université Paris Ouest Nanterre La Défense – CNRS
Nanterre, 92001, France

ABSTRACT

Many organizations are in charge of global security management. This paper outlines and argues for the construction of a theoretical and methodological framework in order to critically assess the new technopolitics currently being developed in the field of global security and which are materialized in standards. The main purpose is to design both a methodology and specific text mining tools to investigate these standards. These tools will be implemented in a platform designed to provide cartographic representations of standards and to assist the navigation of an end-user through a corpus of standards.

Keywords. Standardization, text mining, ontologies

1. INTRODUCTION AND CONTEXT OF THE RESEARCH

The research developed in this paper relates to the industrial standardization of security and risks, located at the intersection of global security and international standardization. The technical and political aspects of these two areas give rise to technopolitics, which we contend will play a leading role in the field of industrial regulation in the XXIst century, causing major industrial, economic and geopolitical impacts.

Many organizations are in charge of global security management. To address this issue, ISO (International Standards Organization), the main international organization for technical standardization, has launched a set of studies in the wake of the ANSI program "Homeland Security Standard Panel" (law of 2002, November 25), subsequently adopted by European and national standardization organizations.

This paper outlines and argues for the construction of a theoretical and methodological framework in order to critically assess the new technopolitics currently being developed in the field of global security and which are materialized in standards. The main purpose is to design both a methodology and specific text mining tools to investigate standards. These tools are designed as heuristics which enable reformulations, semiotic transfer (texts to diagrams), and comparison between texts.

This work is part of a scientific research project called "NOTSEG"[1] which stands for "Standardization and Global Security, The formulation in Standardization of the global security concept". The partnership includes academic and industrial partners: two research centres, MoDyCo from the University of Paris Ouest Nanterre La Défense - CNRS, CQP2I from the Technological University of Compiègne, AFNOR, the French national standardization body, and the company Sector SA, specialized in decision-making and studies in the area of technological and organisational risks.

The NOTSEG project mission is to draw up the cartography of existing normalization frameworks in the field of security and crisis management. The project comprises several successive stages, namely: i) inventorying the list of standards to be studied; ii) analysing existing standardization frameworks; iii) studying and selecting variables; iv) analysing and monitoring a corpus of standards using text mining tools. The methodology of these tools is the topic of this paper.

Recent evolutions in standards are the core of our analysis. The processes of international industrial standardization apply not only to the artefacts of technical devices (in technical standards), but also, since the year 2000, to organizational methods and the evaluation process of these devices, including state regulations, especially in standards of management. Domains such as "business continuity" and "resilience" now possess standards of 'security management".

The rest of this paper is organised as follows. Section 2 explains the theoretical approach and issues. Section 3 details our hypothesis and methodological approach. Section 4 describes the experimentation in progress, section 5 details the platform specifications and finally section 6 presents the conclusion.

2. THEORETICAL APPROACH AND ISSUES

Texts on industrial standards are here considered as means of validating and communicating technical choices, knowledge, and professional practices, as manifested in their cultural and industrial contexts within the wider international and economic situation. We aim to identify, in a specific sector of activity, the conditions and mechanisms that are conducive to new standards contributing to focusing and communicating certain practices, processes, modes of organization and socio-technical arrangements, on the European or international level. In the second step, it will be necessary to investigate whether these new standards really play a role in the evolution of the above-mentioned elements. If this is the case, changes in stakeholders' knowledge and professional practices will then be considered.

[1] The NOTSEG Project (www.notseg.fr) is funded by the French National Research Agency (ANR) over a 3-year period (2009-2012).

In international standardization, the process of establishing each standard is dependent on the conditions of production goals and modalities of use anticipated by the standard. The development of standards implies the analysis, case by case, of the means of action and coordination of activities by organizations and communities [1].

In what way do standards contribute to the communication of knowledge in a specific sector of activity? Standards encode knowledge, modes of organization (such as the « Plan-Do-Check-Act model », a widely used corporate model that structures the management process), approaches (to the nature and evaluation of risks, for instance), or procedures that need rethinking.

In the writing process of standards, our aim is to investigate, beyond the institutional display of a "consensual" operation, a pragmatic dimension in a given sector of activity (its technical committee), the goals which govern their construction, the categories of stakeholders involved, the motivations which influence choices (in terms of writing the contents), the participants in workgroups, and the adjustments that take place between stakeholders from different cultures and languages. We wish to point out that these stakeholders have professional practices in the same domain which can be similar, different, complementary but seldom conflicting.

What conceptual and organizational approaches will be foregrounded (acknowledged) in international standards? Will the different concepts and national practices combine to form new standards? This seems highly unlikely as concepts and practices differ, making them to a large extent incompatible. It can be assumed that the standards which are the most widely acknowledged and implemented on the international or national level will be used to develop the international standard in the domain in question. In the development process, competition may arise between influential national standards and standards backed by lobbies (such as ASIS in business continuity activity, for instance). For this reason, following the example of technical standards, corporate standards, called "management standards" are used by consultants in the domain and by companies aiming to apply for certification of their activities.

Lastly, the intertextuality and performative dimension of standards, which play a crucial role in their application, remain largely unexplored.

The formatting constraints imposed by standards create the conditions of a « performative » [2] that is to say a form of action is present, both intrinsically and extrinsically, in the writing process of these formats.

Hence the importance of:

- monitoring standards during the writing process;
- perusing and comparing standards closely related with the domain under the process of standardisation within different technical committees;
- confronting and attempting to ensure agreement on the definition of terms and concepts.

These three issues are generic and trans-sectorial. In the NOTSEG project, we have chosen BCA (Business Continuity Activity) as the field of experimentation for our methodology but the tools designed are expected to be applicable to other fields.

3. HYPOTHESIS AND METHODOLOGICAL APPROACH

Theoretically, research in progress deals with the relation between media flow of knowledge and natural language processing, in particular text mining tools.

The qualitative analysis of textual data enables data to be processed in such a way as to reveal the heterogeneity of large text corpora. Information traceability places communities (in terms of professional cultures, lobbies, etc.) and memory at the heart of knowledge flow, thus highlighting the interest of a socio-cognitive and political approach to these issues. Grasping and elucidating the diversity of points of view (political, economical, institutional, national, etc.) of stakeholders, who express their opinions in various spatio-temporal and cultural registers, is a tremendous task.

Our proposal is to combine natural language processing and linguistics with the sociology of science, in order to produce a fruitful analysis of corpora [3], since the heterogeneity of the texts imposes complex computing and epistemological processing.

Using concepts from text linguistics [4] is mandatory if one is to identify enunciative polyphony and to build classifications or cartographies. When combined with insights from the sociology of science [5] concerning stakeholders, their institutions and industrial or scientific policies, these representations could provide relevant categorizations of knowledge flows or controversies[2].

In the context of the NOTSEG project, we focus on the role of text mining tools which can provide a new way to apprehend the complexity involved in comparing large numbers of texts. This kind of approach combines linguistic engineering, knowledge engineering and knowledge communication [6].

In the first stage, we aim to check a methodological hypothesis concerning the comprehensive analysis of a corpus of standards by using qualitative text mining tools. More precisely, we attempt to identify the diversity of stakeholders and their institutions by analysing the tracks they have left in the text.

Characterizing stakeholders profiles.

Presently, we have identified several profiles of stakeholders who are likely to provide, prescribe or use standards. A diverse range of practitioners is involved:

- editors of standards: their role is to edit standards by comparing them with other existing standards on closely related topics produced by other NGOs or national, European, or international bodies.
- prescribers and their representatives (companies): these are officials (employees or consultants) in charge of making use of a standard in organizations (public or private) and monitoring the impact of these standards on legislation and regulations.
- stakeholders involved in making use of these documents and in the development of the corresponding operations.
- end users and their representatives.

[2] Massachusetts Institute of Technology, « Mapping Controversies »,[www.demoscience.org/resources/index.php]

Socio-organizational modelling of actors

This modelling step aims at making the connection between standards and their socio-organizational context based on criteria related to geographical and institutional origins and on working group participants. This categorization has been implemented thanks to the work of AFNOR which carried out a cartography of the people involved.

Institutions include international standardization bodies, European or national standardization bodies, representatives of States (Ministries), NGOs, private actors (from industry and French or European think tanks). Participants in working groups - the editor and expert-members - are also taken into account, as are editors and expert-members belonging to companies. Figure 1 shows the different elements of this modelling in the form of a UML diagram of classes.

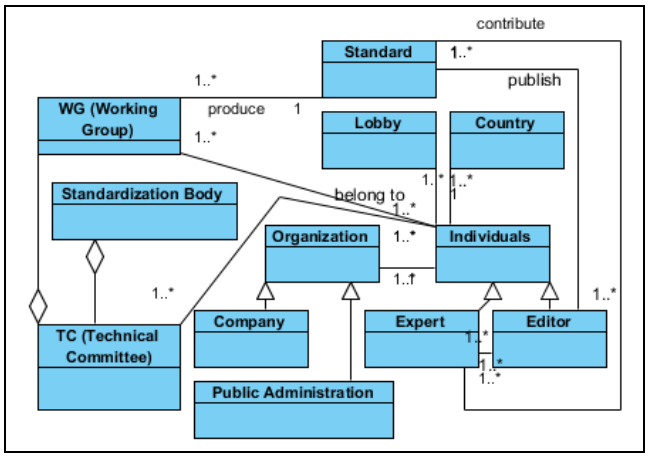


Figure 1: Draft representation of the textual, conceptual and semantic context (UML formalism)

Construction of lexical maps : contrastive glossary of lexical or semantic variations

Undertaking semantic and pragmatic analysis should make it possible to build a dynamic glossary (in English and French) in the field of security – limited first to a sub-field of security which accounts for semantic variations of terms and variations in their uses. This glossary is composed of a list of terms systematically defined in standards in the section "terms and definitions".

This work does not aim at imposing a particular point of view or a unified vision of the field investigated. Conducted according to a comparative approach, this semantic and pragmatic analysis of key-terms in the studied standards aims at revealing zones of convergence, divergence, and cross-checking, including controversies, based on criteria which have to be identified. This glossary should account for semantic variations in terms and variations in their contexts of use. To achieve this aim, several possible methods of work are being investigated:

- the hypothesis of a semantic corpus search with the help of text analysis software (for instance, the list of terms defined in the section "terms and definitions" of the standard).
- expert evaluation concerning the identification of key-terms (corresponding to an interpretative choice of terms),
- a combination of the two approaches.

Concerning the semantic search hypothesis, the techniques and practices related to "clusterisation" (linked to the aggregates method, cf. figure 3) and to graphic – particularly cartographic – representations which derive from it, could enable the construction of networks of categories and concepts, or networks of co-citations to identify actor networks.

Centered on the comprehensive analysis of textual data, computerized qualitative analysis tools can account for the studied corpus in a detailed way and yield a representation in the form of a graph (through tools such as Alceste, Lexico3, or WordMapper, for instance). The construction of networks of associated words (cf. figure 2) is among the techniques implemented by these tools.

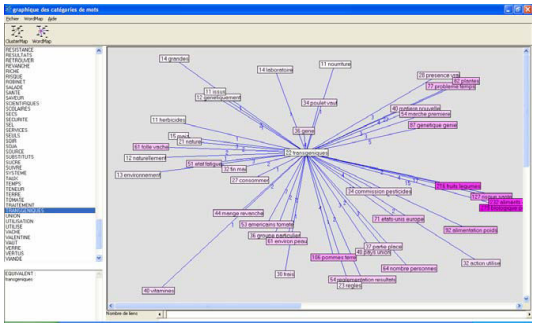


Figure 2: Display of the context of use for a word in WordMapper

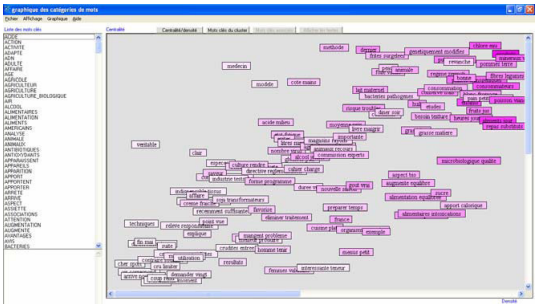


Figure 3: Diagram of clusters in WordMapper

In national, European and international standards, the principle is to analyze documents by comparing language expressions at the level of words, nominal groups, phrases, and paragraphs. The objective is to understand the meaning of identical language expressions in the international standard and in one or several national or European standards, or, on the contrary, their absence or their relative scarcity.

This work is an attempt to go beyond a debate of opinions and to identify a method which provides metrics for comparing texts, i.e. which makes use of statistical results calculated on the basis of the presence of linguistic markers. In other words, it is possible to conduct an analysis in terms of frequency of occurrence or to display the terms which frequently co-occur in a given standard, for instance, thereby displaying the context of use of a key-word and describing its semantic environment. Another issue is to calculate, in a standard, clusters which are built according to indexes of centrality and density. We hypothesize that this cartography would reveal, for a given standard, the semantic environment of terms considered as key-terms (with the option of returning to the text from the diagram clusters in order to carry out a more in-depth analysis).

Construction of local ontologies in the form of graphic and interactive local ontologies

First, we propose to put at the user's disposal tools which offer a synthesis of the standard's contents in a cartographic or textual form. These new representations are designed to help the appropriation of concepts, ways of thinking, ways of organizing or defining modes of operation, these elements being connected to their socio-organizational context.

This work will lead to the construction of a plurality of ontologies – each of them representing a point of view – which will take into account the main "local ontologies" of national, European and international standards.

The originality of this approach is the construction of local ontologies which represent points of view corresponding to semantic worlds of standards studied individually. While an ontology is often used to produce a unified representation of a field, here it is used to express the point of view of a category of stakeholders and a sector of activity.

The methodology consists in constructing a conceptual and lexical representation for each standard, described for example in terms of concepts, organization, practices and economic competition. This work implies the design of a socio-technical device (section 5) with the following characteristics:

- This device, which makes it possible to group or contrast concepts, takes the form of an ontology.
- It is a formal language for the description of concepts and their inter-relations. This ontology is built according to a pragmatic approach, both linguistic and non linguistic, introducing parameters such as actors, editors, committees, nationalities or contexts (industrial or cultural).
- This technical device uses the semantic variations present in terms associated to texts.

Although an ontology is always a more or less objective representation, the construction of an ontology is an essential step in building a conceptual representation. An ontology describes a domain with concepts which are theoretical schemata which favour the intelligibility of phenomena.

Comparative analysis of local ontologies

Then, a comparative analysis of the network of concepts built will be conducted in order to build bridges between them. This analysis will rely on a device which allows navigation between different linguistic expressions of the same concept in all the texts. Some visual support will be provided, with a set of colours for instance, to assist the user.

The goal is to launch the construction of interactive cartographies which facilitate cross-comprehension and provide comparative information about texts of standards and their associated context.
These cartographies will cast light on overlapping, similar, different or convergent text areas between national or international standards.

Another goal is to construct a general ontology from these local ontologies. The ability to identify an abstract concept related to local concepts and their linguistic descriptions in texts of standards would be very useful for monitoring standards.

4. EXPERIMENTATION

In order to assess our methodology, we have undertaken experimentation in the field of BCA. This experimentation comprises the following steps :

1. Corpus selection: selection of standards in a more restricted field than that of information security. We have selected BCA;
2. socio-organizational inventory of the universe of standardization, in connection with national bodies (AFNOR);
3. Building a glossary: Comparative analysis of texts of standards;
4. Building the conceptual map and local ontologies; setting up a referential of terms with the help of experts; identifying words which are important for the description of professional knowledge or practices;
5. Specification of the platform.

In this paper, only steps 1 and 5 are described.

Corpus selection
Within the large corpus of standards on global security management, we have chosen the topic of BCA and risk management because we are involved in Working Group 4 of the ISO Technical Committee TC 223 on societal security. Among the various standards that have been published in different countries (USA, UK, Australia, France), the following have been selected for the present study:

Standard reference	Title
BS 25999-1:2006	Code of practice for business continuity management
BS 25999-2:2007	Specification for business continuity management
ASIS SPC.1-2009	Organizational Resilience : Security, Preparedness, and Continuity Management Systems – Requirements with Guidance For Use
NFPA 1600. 2010	Standard on Disaster/Emergency Management and Programs.
BCI Good Practice Guidelines 2010	A Management Guide to Implementing Global Good Practice in Business Continuity Management
AS/NZS 5050:2010	Business continuity – Managing disruption-related risks, Standards Australia
ISO FCD 22301:2010	Societal security
ISO WD 22399:2010	Guideline for incident preparedness and operational continuity management,

ISO/CEI FDIS 27031	Information technology -- Security techniques -- Specification for ICT Readiness for Business Continuity (FDIS) SC27
ISO/CEI 24762:2008	Technical securities
NF ISO 31000 (ISO/IEC Guide 73)	Risk management - Vocabulary. TMB
IEC/ISO 31010	Risk management – Risk assessment techniques
Draft ISO/IEC	Guide 81 -- Guidelines for the inclusion of security aspects in standards.
NF ISO/IEC IS 27005	Information technology -- Security techniques -- Information security risk management. SC27

5. PLATFORM SPECIFICATION

One of goals of the NOTSEG project is to specify a platform dedicated to the management of a large corpus of standards, between 15 and 100 texts, applied to a engineering field, such as for instance BCA. This platform, designed to provide a common environment, will be used by two kinds of end-users: first, by consultants and editors of standards to help them during the writing process of a new standard (cf. section 3.1); second, by corporate departments in charge of tackling the implementation of standards in order to comply with national or international regulations (cf. 3.2). Two main ideas underlie our approach: one is to provide cartographic representations of standards, the other to assist the navigation of an end-user through the corpus of standards.

Considering that there is no universal representation independent of the goals and the organizational context, we aim to provide both graphic and textual representations, and several tools enabling comparison between several standards. It must be emphasized that all the representations are interconnected and that the platform provides specific interfaces allowing the end-user to navigate between them. Furthermore, this navigation will be assisted by applying specific knowledge based on the NaviText model [7].

Textual and Graphic Representation

The glossary of the standardized domain is the main textual tool. For each term, semantic and usage variations in the selected corpus are provided and enriched links (see section 5.2 below) can be followed; at any moment, the textual contexts of the same term in two standards can be compared by accessing them in one or several standards. This very simple tool is extremely useful to preserve conceptual coherence during the writing process of a new standard by using the same word to refer to an identical concept or on the contrary by choosing a new word to highlight the creation of a new concept.

Graphic representations complete the glossary. As explained in section 3.5, we consider that conceptual maps (or local ontologies) provide a useful level of abstraction, while at the same time keeping and foregrounding the relations between concepts and qualifying their semantics.

For example, relations could be linked to the different phases of the PDCA cycle which governs all the standards.

Assisted Navigation

One of drawbacks of printed standards is that few tools exist for navigating through them, such as indexes and tables of contents. Thanks to digitization, a wide range of possibilities are presently available. The main point is the granularity of the objects accessed. In the NOTSEG platform, the basic granularity is the word. From a structural point of view, words are included in one or more sentences, which are included in a paragraph, which are included in a section and so on. From a text point of view, the frequency and context of a word, and co-occurrence networks, can be computed and linked. From a semantic point of view, the definition of a term in the standard, and the semantics of verbs which co-occur with specific terms, can be automatically annotated with metadata. Finally, from a pragmatic point of view, organizational metadata can be manually added in order to highlight the influence of certain lobbies on the definition of concepts.

All these data and metadata will be used by navigational knowledge [8] to afford different means of circulating through the standards. This kind of tool should enhance the intelligibility and comprehension of standards for engineers responsible for implementing them in companies.

Technical issues

The platform must be interoperable and in line with international standards of knowledge representation (RDF, OWL), and offer API in order to cooperate with other software. Two paths are checked. First, Protegé designed by Stanford University (http://protege.stanford.edu/doc/) is used worldwide and provides a sturdy background to implement the ontology. Furthermore, an endpoint SPARQL is included to store data and metadata.

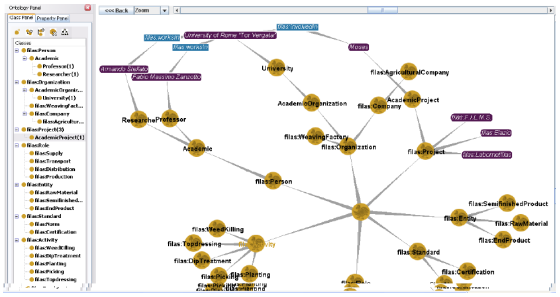


Figure 4: A screenshot from Semantic Turkey [9]

A second option is Semantic Turkey (ST) designed by Tore Vergata University [9], which is a "a Firefox based Knowledge Management and Acquisition Platform for the Semantic Web". The main asset of ST is its ability to combine texts and several ontologies and to provide tools to keep track of concepts and their usages in different texts (see figure 4). Furthermore, an experimentation in the legal domain is in progress at the Artificial Intelligence Department of Roma 2.

6. CONCLUSION

We have described a methodology based on hypotheses of how to combine natural language processing tools with a sociology of science approach.

We have presented the specifications of a platform which is dedicated to editors, prescribers or end-users in the field of standardization and work in progress in the domain of BCA.

This specific study on the texts of BCA standards will be used for a more generic task, as the empirical core of a wider-ranging, prospectively designed inquiry.

The first step is to identify the main concepts and recommendations (operational, behavioral, technical, etc.) in which these formats are embodied. In the second step, this work on texts, related to the socio-organizational context of their production and communication, should provide information to identify or critically assess the new technopolitics already developed or under development in the field of security and crisis management.

Finally, this work will provide the opportunity to begin exploring, in the industrial domain, ways of thinking about the culture of security and risks which are presently the domain of engineering.

7. REFERENCES

[1] L. Thévenot., "Un gouvernement par les normes. Pratiques et politiques des formations d'information", in Conein B., Thévenot L. (eds), "Cognition et information en société", **Raisons Pratiques**, No 8, 1997.

[2] J. L Austin., **Quand dire, c'est faire** (original title: **How to Do Things with Words**), 1962, trad. fr. 1970, Paris, Seuil, coll. Points essais, 1991

[3] D. Demazière., C. Brossard, P. Trabal, K. van Meter (eds), **Analyses textuelles en sociologie, logiciels, méthodes, usages**, PUR, 2006.

[4] J.-M Adam., **Les textes, types et prototypes : récit, description, argumentation, explication et dialogue**, Paris, Nathan, 1992.

[5] B. Latour **Changer de société. Refaire de la sociologie** (Titre original : Re-assembling The Social. An Introduction To Actor-Network Theory, 2005), Paris, Ed. La Découverte, 2006.

[6] B. Juanals B., **La circulation médiatique des savoirs dans les sociétés contemporaines**, Université Paris Diderot-Paris 7, 2008.

[7] J. Couto J., J.–L Minel., "NaviTexte, a Text Navigation Tool" , **Lecture Notes in Artificial Intelligence 4733**, Springer–Verlag, p. 251–259, 2007.

[8] J. Couto J., J.–L Minel., "A Linguistic and Navigational Knowledge Approach to Text navigation" , **IJCNLP,** India, 2008.

[9] M.T. Pazienza, S. Sguera, A. Stellato; "Let`s talk about our "being": A linguistic-based ontology framework for coordinating agents". (R. Ferrario, & L. Prévot, Eds.) **Applied Ontology, special issue on Formal Ontologies for Communicating Agents** , 2 (3-4), 305-332, 2007.

Genetic Traffic Intelligent System

.

Rustom Mamlook, Abdullah Aljumah
Department of Computer Engineering
College of Computer Engineering & Sciences
Al-Kharj University - Saudi Arabia
rstmamlk@hotmail.com

Hassan Mohammed, Malik Amayreh
Department of Information Systems & Computer Science
Amman Al-Ahlya University, Jordan

Issa Abdulhakim
Department of Computer Science-Zarka Private University

ABSTRACT

We considered the application of Genetic Traffic Intelligent System (GTIS), which can be uploaded on the internet and constructed in cooperation with Traffic Engineering Department in Amman city in Jordan. Our system enables the travelers and traffic engineering department to be better informed, and make better traffic intelligent decisions. Our system designed to make the movement of people and goods more efficient and economical. We designed the overall architecture of the system, and also showed the algorithm used to find the optimal path. The GTIS can be used as a guide for traffic navigation by the internet which will be a very useful tool in the next generation of cars and a part of the traffic control systems of the Traffic Engineering Department. Therefore our system could be a base for a very powerful Intelligent System which managing efficiently dynamic geographic and transportation data.

Keywords: traffic, transport, Genetic Traffic Intelligent System, optimal path.

INTRODUCTION

Roads and traffic lie at the heart of modern civilization[1], the problem of controlling the road was and will still be a very important issue, and has been gaining interests to support more efficient control of transportation, since the road capacity is relatively scarce[2-8]. Any traffic system must handle the following issues:

- Efficient control of roads and traffic [1].
- Saving human lives and time [9].
- Augmenting the overall safety of our roads [9].
- Reducing journey times and journey-related trip planning.
- Reducing some of the harmful transportation effects on the environment [10].

Our proposed system can be considered as a navigation devices and traffic information services that inform drivers about the most efficient routes to use in order to avoid traffic delays and wrong directions.

GTIS involves the Geographic Information System (GIS)[4] and analytical/decision models to produce systems able to cope with the problems of controlling the roads. It aims to support decision making by applying quantitative approaches on the geographic information which is stored in a manipulability form within the GIS.

GTIS manages the optimal path that is composed of links and nodes [11], it supports efficient and fast handling of large data for traffic status incident information query, optimal path finding, and many other important features.

GTIS also handles the road data that are composed of links and nodes such as traffic volume, road light, road width, road length, traffic light, road name, road setting, road–blocks, accidents, snow and flooding, and many other factors[9] .

Our system is designed for the city of Amman where its Traffic Engineering Department strived to build safe and effective transportation systems, and employ it in operating and controlling, and can be implemented to any other city with some modifications.

SYSTEM METHODOLOGY

The major targets of GTIS are the construction of traffic information service center, integral traffic information guide, genetic optimal path finding, travel time guide, car navigation support, and all the mentioned goals such as reducing traveling time, minimizing the pollution of environment, guiding travelers through unknown cities and areas, increasing the safety on the roads.

Our system performs the major input/output, presentation, search, and other functions related to the traffic and transport data.

OUR OVERALL SYSTEM

The overall architecture of the Genetic Traffic Intelligent System (GTIS) is shown in Figure 1, GTIS system is consist of four sections, which are[1-3]:

- Integrated genetic optimal path processing section to manage optimal path algorithm.
- General traffic information section to perform the static data related with nodes, roads, and distances.

- Specific traffic information section to integrate traffic information guide.
- Service information section to integrate service information guide.

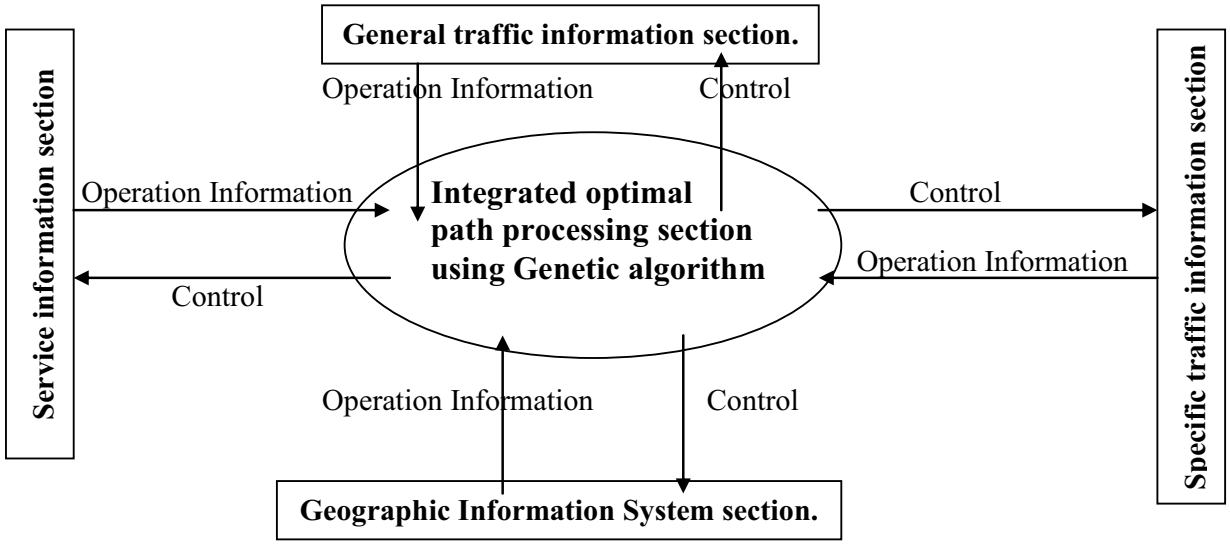


Figure 1. The overall architecture of the GTIS system

INTEGRATED GENETIC OPTIMAL PATH PROCESSING SECTION [12-20]

Intelligent Traffic System using genetic algorithm has been constructed to find the optimal path between nodes and positions, the alternative solution(s) and to suggest optimal path to users. This part is implemented by shortest path algorithm that depends on the shortest distance between links and nodes as the main role to manage the shortest path algorithm. But to manage the optimal path, there are number of factors, such as minimum travel time, average speed, road setting, traffic volume, road light, road width, road length, traffic light, accidents, road-blocks, snow, flooding and the other main factors.

The routing algorithm is widely used in many forms because it is simple and easy to understand. The idea is to build a graph of the subnet, with each node of the graph representing a router and each arc of the graph representing a communication line (often called a link). To choose a route between a given pair of routers[7-8], the algorithm finds the shortest path between them on the graph.

One way of measuring path length is the number of hops. Another metric is the geographic distance in kilometers(Miles), however, many other metrics are also possible but the physical distance is the best one. In the most general case, the labels on the arcs could be computed as a function of the distance and communication cost. By changing the weighting function, the algorithm would then compute the "shortest" path measured according to the physical distance criteria.

When the GENETIC ALGORITHM(GA) is implemented it is usually done in a manner that involves the following cycle:

1. Evaluate the FITNESS of all of the INDIVIDUALs in the POPULATION.
2. Fitness-proportionate REPRODUCTION.
3. Create a new population by performing CROSSOVER operations.
4. MUTATION, which occurs infrequently.
5. Discard the old population and iterate using the new population.

An Iteration of the previous loop is referred to as a GENERATION. The first GENERATION (generation 0) of this process operates on a POPULATION of randomly generated INDIVIDUALs. From there on, the genetic operations, in concert with the FITNESS measure, operate to improve the population.

Given a way or method of *encoding* solution of a problem into the form of chromosomes and given an *evaluation function* that returns a measurement of the cost value of any chromosome in the context of the problem, a GA consists of the following steps (see Figure 2):

Step 1: Initialize a population of chromosomes.

Step 2: Evaluate each chromosome in the population; using a fitness *f*(i) which is assigned to each individual in the population, where high numbers denote good fit.

Step 3: Reproduction: The reproduction (parent selection) process is conducted by spinning a simulated biased roulette wheel whose slots have different sizes proportional to the fitness value of the individuals. This technique is called *roulette-wheel parent selection*.

Step 4: Crossover: Creating new chromosomes by mating current chromosomes, with a probability P*c*.

Step 5: Mutation: Is applied to get new bits in the population, with a low probability P*m*.

Step 6: If the stopping criterion is satisfied, then stop and return the best chromosome; otherwise, go to step 2.

Several algorithms for computing the shortest path are known. Our system is using genetic algorithm, where each node is labeled with its distance from the source node along the best known path. To manage the optimal path, there are number of factors such as minimum travel time, average speed, road setting, traffic volume, road light, road width, road length, traffic light, accidents, road-blocks, snow and flooding and other important factors. For example for the city of Amman :

- If the time is between 7:30-9:30 am or 2:00-4:00 pm then the roads are busy.
- If there is a rain or snow then the roads may be slippery or dangers.
- If the number of traffic lights on a road is more than 4 then the travel time increases.
- If there are bridges, tunnels and traffic lights then the road is safe.

GENERAL TRAFFIC INFORMATION (GTI) SECTION

This section deals with static data related to links, nodes and roads connected between nodes and their cost (i.e. distance), and it supplies data to the optimal path processing section to begin process, as shown in Figure 3.

SPECIFIC TRAFFIC INFORMATION SECTION

It performs the traffic information queries, maps, images and traffic information's display, according to the results of the optimal path processing, and presents it to the user to choose the suitable path because the optimal solution that's produced by the algorithm might not be favored to the user. So the user can use one of the alternative solutions according to traffic information that help him to make decision.

SERVICE INFORMATION SECTION

It performs the information queries, images for the most important service (such as hospitals, market stores, banks, schools and many other services) and shows the services according to the results of the optimal path processing that been chosen by the user.

OUR PROJECT GENERAL INFORMATION SYSTEM (GIS) FEATURES

GIS system has close relationships with the other parts of GTIS project. Major targets are offering convenient and powerful functions to operators and managers to handle input, output, map display and search for the traffic information. So we implemented every useful and diverse function in transparent manner with other information in Oracle.

GIS system consists of Oracle DBMS that manages the data in the system, which is very powerful to manipulate data. Because it performs the searching and analyses very fast, so it is suitable for the system. We carefully designed the database, because it is strongly related to the performance of the overall system as can be seen in Figure 4.

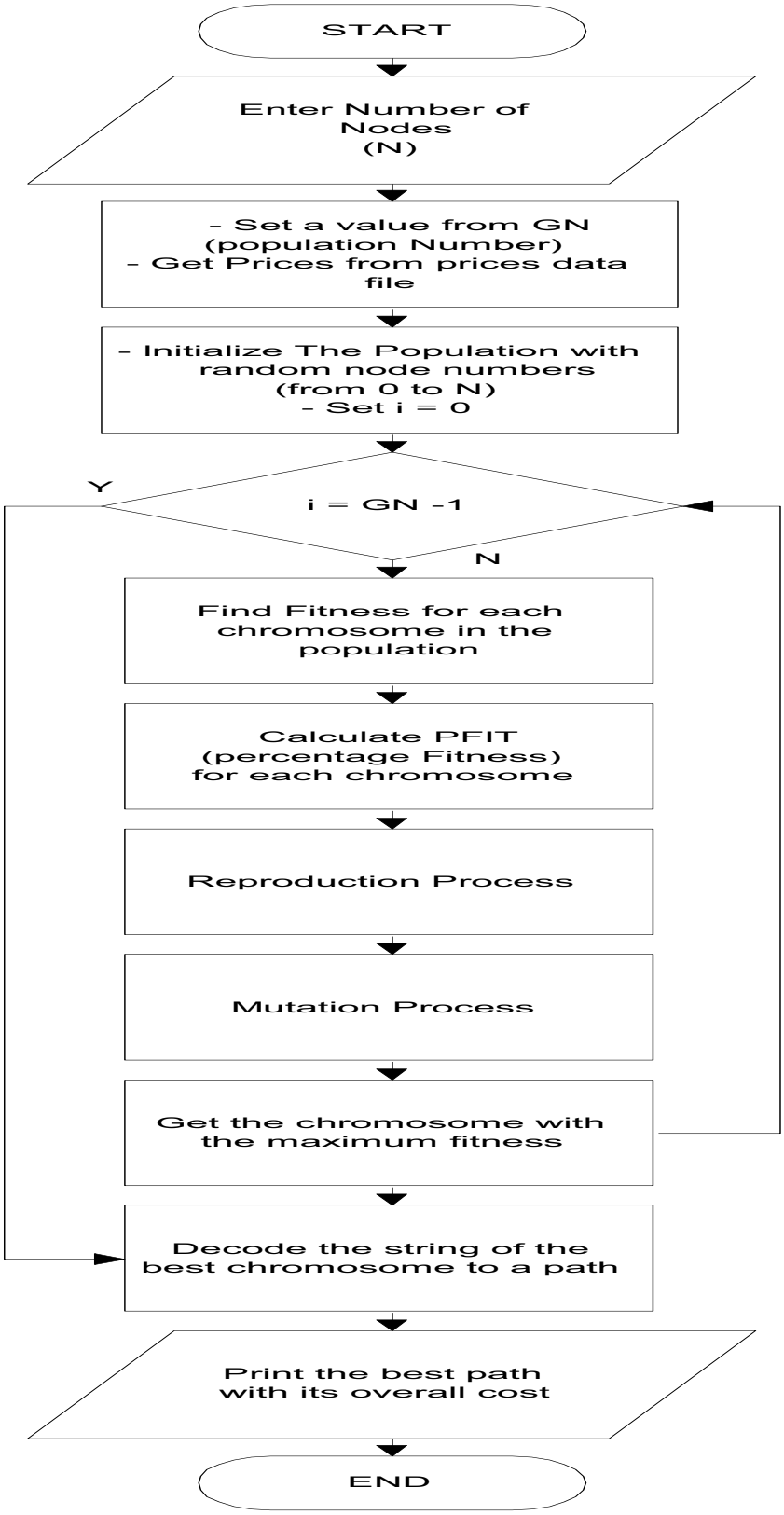


Figure 2 : The *Flowchart* For Short Path Routing Algorithm using Genetic Algorithm

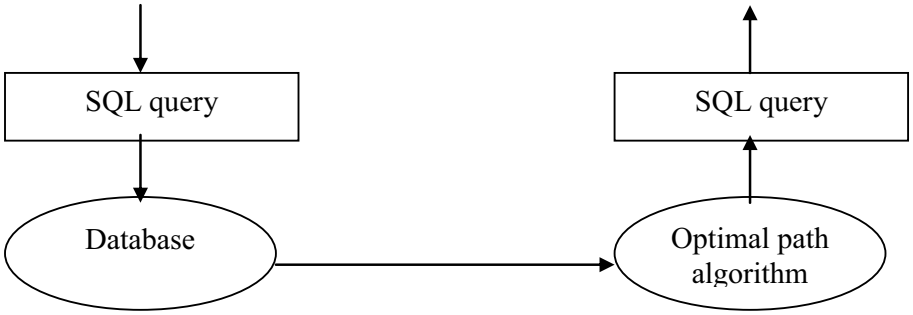


Figure 3. the general traffic information part

Internet Application : In our system, internet is the most important way to deliver the information to end-users. To support interactive application in web, we choose java because it is more suitable and portable in internet than other internet programming languages.

We implemented search, special information listing, and optimal path finding in the nternet service. It's very interactive and friendly to users, and shows good performances and effective solutions, as can be seen in Figures 5 and 6.

Human Computer Interaction (HCI): The prime focus of the (HCI) is how can the user make best use of the system to improve the safety, comfort, efficiency, and productivity of display screen work? This section shows how our understanding of (HCI) has shaped our design philosophy and how we have applied it by using some HCI standards and concepts such as list of values, text editing (editor), system navigators, colors, and other standers.

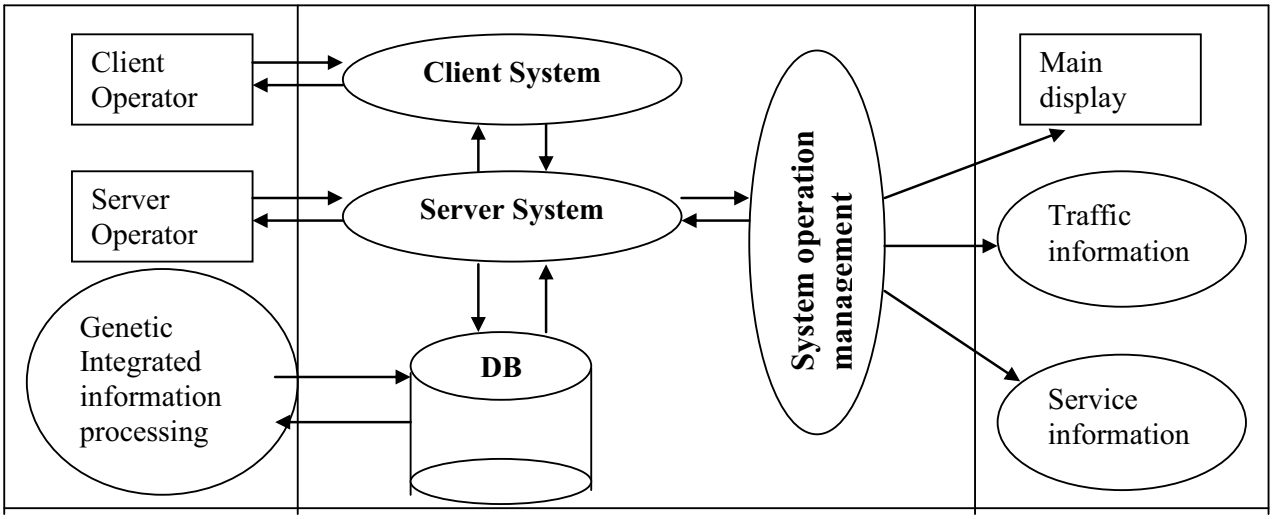


Figure 4.GIS system architecture

Figure 5 .Result of genetic optimal path

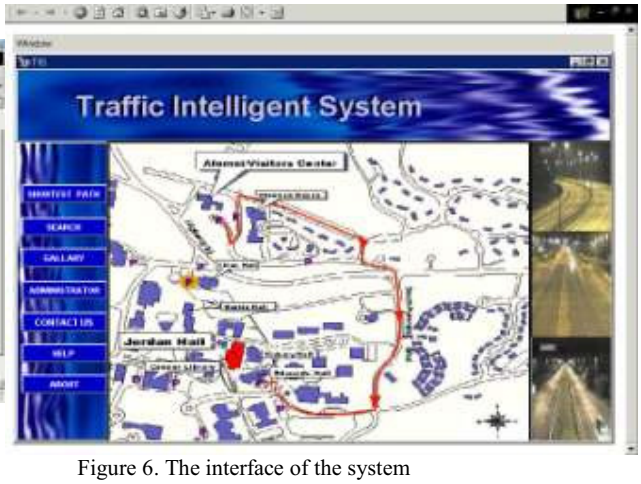


Figure 6. The interface of the system

System Improvements: The system can be improved by adding the following features:

- Using the network to manipulate data of large database in very efficient and fast way.
- Collecting every qualitative and quantitative real time traffic data and static data that is necessary to manage and maintain the real time data, by the sensors and detector and signal controller to deal with a dynamic process.
- Using this system as a base for a very powerful intelligent traffic and transportation system by managing dynamic geographical and transportation data efficiently.

CONCLUTION

The Genetic Traffic Intelligent System (GTIS) can make every journey more comfortable, safe, and less stressful. GTIS helps to find the optimal path between two locations using the gathered data that has been stored in the database. The user can get the information he would like to receive through the system, such as optimal distance, minimum travel time, average speed, road setting, and many other different factors, the different options are calculated using SQL queries on the database and can easily be expanded. The implemented system shows efficient, active, scalable facilities, functions, and many other features in traffic intelligent environment. A freely flowing traffic creates less than slow-moving or stationary traffic**,** so our system which diverts travelers to less congested roads, helps in lowering the level of pollution.

REFERENCES

[1] N.B. Hounsell, and G. Wall, "New Intelligent Transport System Applications in Europe to Improve Bus Services", **Journal of the Transportation Research Board,** Washington DC, Jan., 2002.

[2] N.B. Hounsell, K.Gardner, and D.Cook, "Intelligent Bus Priority in London: Progress on Large-scale Implementation". **Proc. of 9th World Congress on Intelligent Transport Systems**, Chicago, USA, Oct. 2002.

[3] G.Lyons, and I.Chalmers, "The Integration of Highway and Public Transport Information". **Proc. 11th International Conference on Road Transport Information and Control**, March 2002, London.

[4] "Road Traffic Advisor - Achievements and Future Prospects". **Report by Proc. 11th International Conference on Road Transport Information and Control**, March 2002, London.

[5] Metro Investment Report. LADOT's Allyn Rifkin is Not Counting on a Silver-Bullet Fix to LA's Traffic Woes. March 2005.

[6] C. Gershenson, "Self Organising Traffic Lights", *Complex Systems*, Vol. 16, No. 1, 2005, pp.29-53. <http://uk.arxiv.org/abs/nlin/0411066v2> (12 August 2008)

[7] D. Anderson and W. Rose (ed.). ***Transport Trends****: 2007 edition*,Department for Transport, 2007. Crown Copyright. <http://www.dft.gov.uk/pgr/statistics/datatablespublications/trends/current/transporttrends2007> (12 August, 2008).

[8] S. Lämmer, and D. Helbing, "Self-control of traffic lights and vehicle flows in urban road networks", **Journal of Statistical Mechanics: Theory and Experiment,** Vol. (04), 2008, pp.4019-4054. <http://arxiv.org/abs/0802.0403> (14August 2008).

[9] Report by the European transport safety council on "**Intelligent transportation systems and road safety**", thirty-second session 1999.

[10] D.K.Kahaner, **Intelligent transport system the Asian technology information program**, **seminar** Tokyo 1996.

[11] V.K. Sherbo, standard of computer networks. Intercouplings of networks: the Quick reference.- M.:KUDITS-MODE, 2000.

[12] G. Chryssolouris, and V. Subramaniam., "Dynamic scheduling of manufacturing job shops using genetic algorithms." ***Journal of Intelligent Manufacturing***, vol.12, No.3, 2001, pp.281-293.

[13] M. Jensen, "Generating robust and flexible job shop schedules using genetic algorithms." ***IEEE Transactions on Evolutionary Computation***, vol.7, No.3, 2003, pp.275-288.

[14] Haupt. R.L. and S.E. Haupt, 2004. Practical genetic algorithm. John Wiley and Sons, Inc., Hoboken, New Jersey, pp: 27-30.

[15] *Genetic Algorithm* OpenStreetMap.org. <http://www.openstreetmap.org/?lat=51.448987&lon=-0.966427&zoom=18&layers=B00FTFT> (12 August 2008).

[16] *H. Toossian, S., M. Khosrav,i and M. Doaee*, "Intelligent Transport System Based on Genetic Algorithm", **World Applied Sciences Journal** , Vol. 6, No. 7, 2009, pp. 908-913.

[17] H. Kanoh, S. Tsukahara, "Solving Real-World Vehicle Routing Problems with Time Windows using Virus Evolution Strategy", **International Journal of Knowledge-Based and Intelligent Engineering Systems**, accepted (2010).

QUALITY OF LIFE INDEX ASSESSMENT FOR SURAT CITY OF INDIA

Narendra N. Patel
Associate Professor of CED S. V. National Institute of Technology, Surat, Gujarat, India.

Smit N. Patel
Assistant Professor, Vidyabharti Trust Institute of Technology & Research Centre, Umrakh, Gujarat, India.

Dr. Krupesh A. Chauhan
Associate Professor of CED S. V. National Institute of Technology, Surat, Gujarat, India.

ABSTRACT

For many years, economic indicators such as the gross national product and customer price index have been the primary measure on "progress "available to decision makers. However, decision makers are being challenged to produce change that improves the quality of life in social, economic, environment, health, political, education and infrastructural sense. It was also generally said that if a big industrial complex is established anywhere, it will lead to a marked change in the surrounding area in terms of general changes in development and implement in employment opturnities, income generation and consequently in the quality of life of people. But unfortunately small towns have not attracted industries or people and have not served as counter magnets to the growth of big cities; as such there is deterioration in the quality of life of dwellers of big cities due to this very reason.

Also the current patterns of the urban development are based on the ideas imported from western countries and they use systems that are highly capital and natural resources intensives. Capital intensity divided the urban population in to urban rich and urban poor while the resource intensity destroys the rural hinterland. The end result is that while there is an effluent class with a very high quality of life, the majority of urban dwellers face a very poor environment quality.

The urban population in India has increased significantly from 62 million in 1951 to 285 million in 2001 and is estimated to grow to around 540 million by the year 2021.it would be touching 37% of the total population next 15 years. India's urban population has growth rate nearly 41% in last decade. This directly affects the quality of urban life.

In the present paper an attempt has been made to evolve a composite index for quality of life urban dwellers and the same has evaluated for Surat city by conducting pilot survey (100 Nos.).By considering seven main indicators like Natural Environment, Physical Environments, Health, Economic, Social, Political, Educational and Infrastructure affecting quality of life. The subjective measures for all factors given by stake holders and objective measures by expert, to a normalized scale. The importance weighing, on a rank order scale. The study offers a way of expressing the quality of life index using mathematical modelling. The indices of seven indicators that were used to complete QOL Index. The quality of life index for Surat city is 5.11.

This study also helps local Government, policy makers, planners, non government organization and research scholars for decision making tool for development of city region.

Keyword: Quality of life, Urban development, Environment, Infrastructure, Health, Education.

1 INTRODUCTION

Migration of population to urban centres, particularly to the industrial cities, gives impetus to the growth of metropolitan areas. As economics mostly depended upon industry, trade and commerce, these activities concentrated in metropolitan centres, where skill, design and market were readily available. White collar jobs grew faster than factory jobs in view of the fact that more people were needed to design, finance and organise the production and selling of the goods both within the country and outside. [1] These trends caused rapid increase of population in major urban centres.

The increase in urban population has its effect on towns and cities. Over congestion in urban areas, resulting from the increasing exodus of rural population has gradually reduced even healthy and good areas into slums.

Quality of life is emerging as a central construct within many disciplines, such as those comprising the social sciences, economics, and medicine. Its attractions, in part, are that it offers an alternative to some traditional disciplinary views about how to measure success. [2] First it's directs attention onto the positive aspects of people's lives, thus running counter to the deficit

orientation of these disciplines. Second, it extends the traditional objective measures of health, wealth, and social functioning to include subjective perceptions of well-being.

2 STUDY AREA

Surat is known as "The Silk City", of India. It is also known as "The Diamond City", and "The Green City", etc. Surat is one of the cleanest cities of India. It has the most vibrant present and an equally varied heritage of the past. Still today, Surat continues the same tradition as people from all around the country comes for business and jobs. In Surat, jobs are easier to get here due to very fast development of various industries in and around Surat City. The city of Surat is fast growing industrial city and is the 9th largest city of the India as per Jawaharlal Nehru National Urban Renewal Mission (JnNURM). From time to time jurisdictional limits of SMC have also been extended to include the outgrowth. At present SMC area is 326.515 Sq. Km which was 112.28 Sq. Km before the city limit extension in July 2006. The new physical boundary of Surat with respect to Surat urban area is as shown in Fig 2.1[7] There are about 6.50 lakh people (2001) reside in the immediately periphery of the city. Surat is now considered as a Mega City of India.[8] In census year 2011, it came out with the highest growth rate among cities of Gujarat state. Moreover, Surat was the second 'Most growing City' of India as per Census 2011 with a rate of 62.38%. SMC'S population rose from 28.42 lakh to 44.72 lakh between 2001 and 2011.

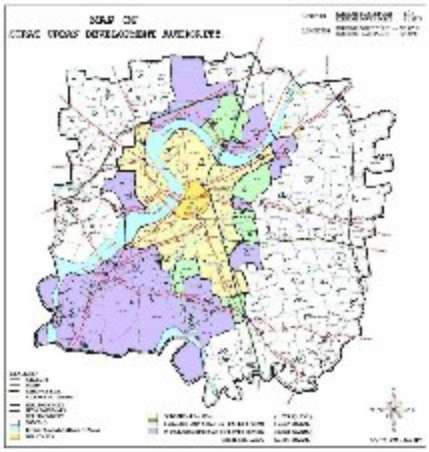


Fig.2.1 Surat city and Surat Urban area.

3 OBJECTIVE

- To assess the Quality of Life Index for Surat city

4 METHODOLOGY

It is felt that the approach of this study should be semi-empirical and that the efforts are made to collect the data from the concern department as well as by directly talking to the local people. The inventory study approach has been adopted. Main aim is assessment QOL Index for Surat region. Each basic services id derived by giving them ranking according to city peoples' opinion. For better study result and time constrain the study area is limit to south-west zone. The peoples' opinions are to be taken by visiting this zone one by one and data are to be collected. After collecting the all data, making Mathematical model for QOL Index measurement [5] Measure the quality of life index for this zone.

5 QOL INDEX MODEL

The evolution of the QOL Index was based on the model developed by Mr. Kenneth E. Hornback. The following values are required to be obtained in order to evolve the life index values.

- S_{ij} - The subjective, or satisfaction measure for all factors given by people to a normalized scale 1-10.
- O_{ij} - The objective measure for all the factors as given by experts also normalized to scale 1-10.
- W_{ij} - The importance weighting which the individual attaches to a particular factor, relative to all the other factors, on a rank order scale.

All the above mentioned three values were obtained from individual and experts who were permanent residents of this city. A special form was prepared based on the work of this study and a sample pilot survey was carried out for this city.[3],[4],[6].

The analysis was done for this forms in the following steps:

I. The sample forms obtained were broadly classified into 5 different groups viz. Graduates, professionals, females, below S.S.C. and others.
II. To obtain the S_{ij} values for different individuals for different factors was an important thing. The individuals were to read the statements corresponding to different factors and were to circle any one of the following code words.
III. The S_{ij} value was to be obtained normalized to 1-10 scale. This was obtained by dividing the statements into two sections – one which would carry higher weightage if a person strongly agreed with it and the second which would carry higher weightage if a person strongly disagreed with the statement.
IV. To obtain the O_{ij} values, different experts in different fields were interviewed personally and

were requested to give weightage normalized to a scale 1-10, depending on the objective indicators for a factor. The experts included ex-mayor, doctors, engineers, professors, etc.

V. The individuals were also told to give weightage to all the factors relatives to all other factors, on a rank order scale. This was value of Wij.

VI. After obtaining the required values the analysis was done with the help of a computer. A computer program was developed and data was fed to the computer to give the QOL Index.

6 DATA AND ANALYSIS

Factor Introduction

The physical input of the overall quality of life consists of seven principal goal areas or QOL Index components. They are defined in broad terms, and cover most major concerns of all individuals :

I. Environmental components
II. Health components
III. Economic components
IV. Social components
V. Political components
VI. Educational components
VII. Infrastructure components.

These concerns have been chosen with a view to developing as broad and common as possible a concept of well being. The Environmental component discusses the Natural Environment and Physical Environment. In this discussion we shall see the importance of each component and we will list out the factors to be considered in each component.

Sample Data for South West Zone

In this study to use of above data the comparison between population and factors, analysis are done and see comparison in to graph. The work is carried out for seven zones of Surat. In this paper the representative data for south west zone has been given in Table No. 6.1 to 6.8 and Graph No.6.1.

	Physical Environment				
Parameters	A	B	C	D	E
Housing	55	41	4	0	0
Transportation	0	19	64	17	0
Public Service	15	49	32	4	0
Aesthetics	20	64	16	0	0

Table No. 6.1 Physical Environment

	Natural Environment				
Parameters	A	B	C	D	E
Air Quality	19	30	30	20	1
Water Quality	28	42	16	14	0
Solid waste	13	42	21	23	1
Noise	29	38	19	14	0

Table No. 6.2 Natural Environment

	Health Component				
Parameters	A	B	C	D	E
Physical Health	12	49	36	3	0
Mental Health	8	37	44	11	0

Table No. 6.3 Health Component

	Economic Component				
Parameters	A	B	C	D	E
Income	13	65	18	3	1
Income Distribution	8	42	43	7	0
Economic Security	9	35	32	22	2
Work Satisfaction	5	71	23	1	0

Table No. 6.4 Economic Component

	Social Component				
Parameters	A	B	C	D	E
Family	15	63	16	6	0
Community	8	56	28	6	2
Social Stability	22	60	18	0	0
Physical Security	22	39	31	6	2
Culture	11	55	23	10	1
Recreation	6	59	33	2	0

Table No. 6.5 Social Component

	Political Parameter				
Parameter	A	B	C	D	E
Political	6	73	19	2	0

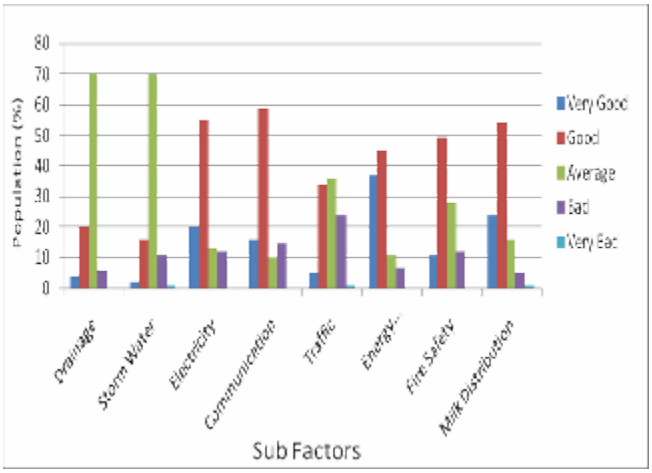
Table No. 6.6 Political Parameter

	Education Parameter				
Parameter	A	B	C	D	E
Education	12	56	21	11	0

Table No. 6.7 Education Parameter

Parameters	A	B	C	D	E
Drainage	4	20	70	6	0
Storm Water	2	16	70	11	1
Electricity	20	55	13	12	0
Communication	16	59	10	15	0
Traffic	5	34	36	24	1
Energy Distribution	37	45	11	7	0
Fire Safety	11	49	28	12	0
Milk Distribution	24	54	16	5	1

Table No. 6.8 Infrastructure Component



Graph No. 6.1.Infrastructure Component

7 QOL INDEX ASSESSMENT

In this study, an attempt has been made to evaluate the 'QOL Index' for Surat region, by making use of the mathematical model developed by Kenneth E Hornback and others. The value of the index is measured on a scale normalized to 1-10. The value of index nearer to 10 indicates an excellent quality of life, whereas the value nearer to 1 indicates the worst quality of life.

QOL Index is an average value of the factor indices (F_j values) of different factors. Thus factor having low F_j value would tend to bring down the value of the overall index of quality of life. Depending on this, the factors can be grouped under various heads like poor factor indices, medium factor indices and better factor indices. In poor factor there are two sub-factor included. In medium factor there are twenty two sub-factor included and in better factor there are six sub-factor included.

Zone	Major Factors (Fj Value)							Average	Remark
	Env.	Health	Economic	Social	Political	Education	Infra.		
West	5.72	5.07	5.48	4.82	5.14	5.76	5.46	5.24	2
S W	5.92	5.23	5.82	5.09	5.28	5.72	5.44	5.28	1
North	5.30	5.13	5.22	4.99	5.37	5.27	5.40	5.17	5
Central	5.10	4.98	5.17	4.92	5.24	5.51	5.31	5.19	3
South	4.83	4.80	4.83	4.71	4.54	5.21	4.96	4.83	7
S E	4.83	5.17	5.21	4.48	5.10	5.02	5.09	4.93	6
East	5.14	5.11	5.06	4.90	5.54	5.37	5.46	5.18	4

Table No. 7.1 Zone wise Factor Index and Rank

8 CONCLUSION

In this study to use of all data the comparisons between population and factors, In this study, an attempt has been made to evaluate the 'QOL Index' for Surat city.

The composite value of the index obtained on analysis is a result of the combination of the following factors: (I) Natural Environment (II) Physical Environment (III) Health (IV) Economic Sector (V) Social Sector and (VI) Political Sector (VII) Education Sector (VIII) Infrastructure Sector.

Surat city in Gujarat has the highest growth rate. This study has indicated that the QOL Index for south west zone having highest rank and south zone is having lowest rank among the seven zone of Surat city. Surat city of India is having QOL Index as 5.11. The industrialization and related urbanization are therefore not the indicators of improvement in the living conditions of the people. It is therefore desirable to use the QOL index of the urban dwellers as a tool to guide the development and in preparation of the master plan of the Surat city.

9 REFERENCES

[1] Bloom, David E., Patricia H. Craig, and Pia N. Malaney (2001), "*The Quality of Life in Rural Asia",* Oxford University Press, Hong Kong

[2] Dasgupta, P., and Martin Weale (1992), "On Measuring the Quality of Life", *World Development*, Vol. 20, No. 1: 119-131.

[3] Drenowski, J. (1974), "On Measuring and Planning the Quality of Life", Institute of Social Studies, The Hague.

[4] Osberg, L., and Andrew Sharpe (1998), "An Index of Economic Well-being for Canada", *Research Paper R-99-3E*, Applied Research Branch, Human Resources Development, Canada, Ottawa, Ontario.

[5] R.A.Christian (1988) "Quality of Life Index of Urban Dwellers – A Case Study for Surat City, Research Work SVNIT, Surat

[6] Slottje, Daniel. (1991), "Measuring the Quality of Life across Countries", the *Review of Economics and Statistics*, Vol. 73: 684-693.

[7] Surat city development plan ,2006-2012, surat municipal corporation and surat urban development authority,may-2006

[8] Surat vision-2020, surat municipal corporation, May-2004

* * *

The World of Nanotechnology and its Applications

R. Challoo, S. Terala, and L. Challoo
Texas A&M University-Kingsville
Kingsville, Texas 78363-8202, U.S.A.

ABSTRACT

Nanotechnology is, broadly speaking, an emerging technology that enables engineers to design and build new materials and products at the molecular level.The impact of nanotechnology is already being felt in the form of new computer memories that provide rapid access to stored data, that can hold more of this data than the minidrives used in iPODs and do not need any external power source to retain the data. The impact is also being felt in the form of prototypes for photovoltaic cells that can literally be sprayed onto buildings or computers to provide cheap power sources. It also is being felt in the form of "nanoengineered" gels that speed the recovery of damaged nerve cells.

Not surprisingly, a technology this powerful is attracting attention. *The Economist*, *BusinessWeek*, and *Red Herring* (a magazine for venture capitalists) have all ran cover stories or special reports on nanotechnology. New books on nanotechnology and the underlying nano*science* are also appearing at an accelerating pace. In this paper, we have presented a detailed study of nanotechnology and its applications along with its advantages and disadvantages.

Keywords: **Nano**-technology, -engineering, -electronics, -photonics, -medicine

1. INTRODUCTION

1.1 Understanding Nanotechnology

Nanotechnology involves research and development of materials or equipment at the atomic and molecular level. A nanometer is a measure equal to one billionth of one meter, occupies about 10 atoms. The formulation of a precise definition of nanotechnology is a difficult task. There is debate as to whether naturally occurring nanoparticles such as carbon soot, fall under the rubric of nanotechnology. Finally, some reserve the term "Nanotechnology" for the industry with atomic precision while others use the term nanomaterials to describe the construction materials, devices and systems. According to the Foresight Institute, a nonprofit organization dedicated to preparing society for nanotechnology, "Molecular nanotechnology will be reached when we are able to build things from the atom, and we will be able to rearrange matter with atomic precision"[1]. The National Science Foundation, on the other hand, defines nanotechnology as "research and technological development at the atomic scale, molecular or macromolecular levels, the length scale of approximately 1 to 100 nanometer range, a fundamental understanding of phenomena and materials at the nanoscale and to create and use structures, devices and systems that new properties and functions because of their small size and / or intermediate size."[2]

1.2 History of Nanotechnology

Most epoch-making technologies have been thought of, or even developed, in some primitive form, long before the epochs that they define. Nanotechnology also has its own history of this kind. Medieval glassmakers sometimes used to color stained glass with the help glass nanoparticles. This is real nanotechnology, because particles of different size provided different colors. Depending on their size, gold nanoparticles can show up in glass as orange, purple, red, or greenish

"The current top-down method for manufacturing involves the construction of parts through methods such as cutting, carving and molding. Using these methods, we have been able to fabricate a remarkable variety of machinery and electronics devices. However, the sizes at which we can make these devices is severely limited by our ability to cut, carve and mold"[4].

Bottom-up manufacturing, on the other hand, would provide components made of single molecules, which are held together by covalent forces that are far stronger than the forces that hold together macro-scale components. Furthermore, the amount of information that could be stored in devices build from the bottom-up would be enormous.Although the limitations of traditional chemistry were criticized, recent advancements have improved its potential uses for nanotechnology.

1.3 The Industrial Structure Giving Rise to Nanotechnology

Many industries are using nanotechnology for their product development .The interesting aspect of the nanotechnology is that it has the applications in every field.
Nanotechnology seems set to provide some major benefits to the pharmaceutical industry in the near-term future, both in terms of drug discovery and in terms of drug delivery. These will enable big pharmaceutical firms to find new blockbuster drugs and add to the life of those already on the market or in the pipeline. Nanotechnology will also help improve diagnostics, but mostly in an evolutionary way, especially by offering ways to enhance labs-on-a-chip and medical imaging. There are many ways in which nanotechnology can lead to a great leap forward for regenerative medicine. These include some very near-term opportunities, such as stronger, safer stints. There are also interesting possibilities for using nano-engineered "trellises" to help grow damaged cells.

2. NANOTECH IN THE SEMICONDUCTOR, COMPUTING AND COMMUNICATIONS INDUSTRY

The semiconductor industry is different from the other sectors, in these other sectors, nanotechnology has a profound impact on the products and services the way they are produced, and the way they operate. When nanotechnology is at work for a decade or more, many sectors will be transformed completely.

"CMOS is hard to beat and the semiconductor industry will not totally abandon it, or conventional ways to make chips for many years, Nano or no Nano tech. Unfortunately, this kind of thinking is often taken to extremes, and it is in no way unusual to encounter people in the semiconductor industry, which examines that there is really no reason to spend much time worrying about completely new ways of doing things. For some it may actually be a sensible position"[2].

2.1 Spin and Nanoelectronics

Roughly spintronics uses the properties of quantum spin. Spin is closely linked with the more familiar feature of magnetism. Use of electron control spin is a favorite among researchers for many years, and almost every large electronics company has toyed with the idea at some time or another. One product that can be regarded as an adopted child of spintronics is giant magnetoresistance (GMR) read / write heads that are already widely used in the disk drives and also the Magnetic RAM (MRAM) is one of the latest developments. "This new form of a low energy, non-volatile "memory is nano" produced or being developed by some of the biggest names in the semiconductor industry. These include Freescale, Hewlett-Packard, Honeywell, IBM, Infineon, NEC, Sony and Toshiba. GMR and MRAM are inherently nano technology because of the thin layers required in the building units" [9]. Some of the largest semiconductor companies are now very interested in spintronics at the next level.

2.2 Nanoelectronics with Molecules

The nanoelectronics platforms that we have discussed so far, although quite revolutionary, bear a striking resemblance to the good old-fashioned silicon microelectronics. Molectronics (sometimes called "moletronics") is different, because it has a lot further to go commercially. In fact, it isn't even all that well-defined. Indeed, molectronics is basically a catchall term for electronics that uses complex (often biological) molecules as the main materials platform. There is a considerable amount of interesting R&D in this field that is being performed in both academic and industrial laboratories that involves quantum computing or chemical computing.

There is another aspect to molectronics, though, one that is one of the recent developments."This approach is one in which switching is based upon the change in state of individual molecules. The change can be based on conductivity as a result of an applied field (much like a classical FET), a conformational change resulting in a change of conductivity (including a complete contact break) or optical properties. In practice, devices may contain many such molecules in an individual switching unit. Although some approaches to this kind of molectronics involve radical changes in circuit design, others are more in line with conventional architectures"[10].

2.3 Plastic Electronics:

Plastic electronics based on thin film transistors (TFT), produced by organic polymer films, gives directions to new electronics, including the creation of a whole new range of products that could not be manufactured using conventional CMOS approaches. The discovery that polymers could be much more conductive using doping is what makes this research field interesting. The team of researchers who discovered this won the Nobel Prize.

Plastic electronics generates very little heat and uses small amounts of power, alleviating major problems that dog conventional electronics. The approaches to plastic electronics with the greatest commercial prospects are the ones that emphasize these advantages to the fullest. This is because plastic electronics cannot match the performance of CMOS circuits, so its competitive advantage must be found in features and capabilities that cannot be matched by CMOS. There are opportunities to create new firms and new revenue streams for older firms from these products.

One of the main new product opportunities that seem to be suggested by plastic electronics is disposable electronics for RFID tags and smartcards, to name a few. Memories using low-capacity but plastic electronics has the capability to lead to a new breed of smart tickets, greeting cards, and other products that may include more sophisticated circuitry and even small displays.

Another important new product direction for plastic electronics is electronic paper. This is a "cool" name for a special kind of display that emulates (literally) the look and feel of real paper. It is thin, flexible, high resolution, and (in some versions) will even feel like paper. The difference is that "e-paper" is still a display, which means that what is being shown on the screen can always be updated electronically, including over a network. E-paper will find applications in a number of areas. There are already electronic book readers and e-paper systems can also be used for easily updatable signage in stores and hotels.

Light-emitting diodes (LEDs) have come a long way from the little red and green flashy things that you find on modems and car dashboards. One direction that they have taken is towards high brightness LEDs (HB-LEDs), based on galliumnitride. These are already widely in used in home and industrial lighting, flashlights, car headlights, and so on. Another type of LED is currently being commercialized that are based on plastics/polymers or on smaller molecules and could be the basis of a wide variety of displays, including e-paper. Such organic LEDs (OLEDs) may also ultimately find their way into lighting systems.

Conventional semiconductor fabs now cost in the billions of dollars and are expected to escalate for the foreseeable future. This ability to customize is associated especially with ink-jet printing and other "maskless" printing technologies that do not require huge fixed costs per run.

The "printable electronics" concept seems likely to generate some new business ideas. It is possible to imagine, although sometime in the distant future, a store, much like today's photocopying store, where circuit designers and entrepreneurs could go to bring their design into plastic realities for test, sampling, or other low-volume requirements. Even if high-speed plastic electronics processing were not available, an engineer or businessperson may want to try out a CMOS concept in plastic in order to get a low-cost answer to certain questions that arise about functionality. This kind of vision may be a little futuristic, but it is possible to imagine plastic electronics manufacturing being used to generate application specific integrated circuits (ASICs).

Plastic electronics is about printing with organic or nonmetallic inks onto flexible substrates. It therefore opens up the business to firms such as Xerox and Hewlett -Packard, who have long histories in the non impact printing business. Some materials firms are also likely to see plastic electronics as an emerging opportunity.

Conductive polymers and flexible substrates that are capable of supporting plastic electronics are commercially available. There is a growing amount of research in this space and no one yet knows what will be the standard materials platforms that will support the plastic electronics of the future.

Products that can be created using plastic electronics, but could not be created using CMOS appear to have a ready market. It

seems obvious that a large roll-up display that can help turn a cell phone into a computer or entertainment device would ultimately find significant demand.

While the current generation of plastic electronics is pitched at markets in which it will not compete with CMOS, no one really knows how and where the two materials/technology platforms will ultimately compete. Plastic electronics still seems to be a long way from providing an alternative to CMOS-based processing and logic. On the other hand, some of the theoretical work that has been done suggests that organic material could be used to create processors up to 1 THz, of course, only if one could find the right material.

2.4 Nanophotonics

The impact of nanotechnology on photonics applications is likely to be quite extensive and very much in tune with the normal development of photonic devices, whose evolution in pre nanotech days have often been based around new materials developments. Thus, lithium niobate (LiN) helped improve optical modulators and amplifiers and indium phosphide (InP) offered the promise of integrated optics and electronics."[15] Now it may be the turn of nanophotonics to add its contribution. A number of start-ups have opened their doors in this space and the European Union has launched a large R&D program focused on this area. The impact of nanophotonics is fairly diverse. NanOpto makes a range of nanoengineered polarizers, splitters, and "waveplates," using a nanoimprint lithography.

3. NANOTECH AND ENERGY

3.1 The Real Energy Crisis

The semiconductor industry may have a crisis pending in its inability to carry the CMOS paradigm forward down the path set for it by Moore's Law. The energy industry has a crisis all its own, however, often presented as a shortage of energy. This portrayal is based primarily on the theory that much of our energy comes from fossil fuels and that we are quickly running out of those fuels. Convincing as this theory may be, it is somewhat of a mischaracterization of the energy industry's real crisis. Presenting the opportunities for nanotechnology in the energy sector as largely defined by supposedly dwindling petroleum reserves is likely to lead to an underestimation of the opportunities and perhaps to misunderstandings about what those opportunities actually are.

There is a lot more involved in the energy field than just energy generation. Energy must be changed into different forms, stored until needed and then transported efficiently. We are primarily concerned with how nanotechnology is changing the economics of power, rather than simply changing the cost of fuel. This translates into a broader range of opportunities for nanotechnology than might have been perceived if a purely fuel based-analysis had been applied.

3.2 The Impact of Nanotechnology on the Energy Sector

As we have seen there seems to be a general agreement that nanotechnology will have a big impact on the future of the energy sector, even if no one can quite agree on what that impact (or what that future) will be. Nonetheless, based on the technology directions currently being taken by both R&D efforts and corporate commercialization programs, it seems reasonable to assume that the impact can be categorized into five reasonably well defined headings: 1. The nano-enhanced fossil fuel sector; 2. Fuel cells and the nanoengineered hydrogen economy; 3. Nanosolar power; 4. The nano-enhanced electricity grid of the future; 5. Nanopower for the pervasive communications network.

The vast majority of nano energy businesses will find that they fit pretty well into the areas listed above, which also overlap each other in some ways. The end game for the nano-enabled energy sector could be a dramatic change in the world energy picture, with major disruptions in the kinds of energy used by industry and consumers and opportunities emerging for those currently without access to reliable energy sources.

3.3 Nanotech, the Environment, and the Road to the Hydrogen Economy

Pollution is a topic that is intimately tied up with the energy sector, because most of the fossil fuels are highly polluting. Nanotechnology firms may well help in this regard using nanocatalysis, and as a near-term opportunity. Nanocrystalline catalysts made from cadmium, selenium, and indium have shown to be effective carbon dioxide filters, while titanium oxide nanocrystals under UV light will remove mercury vapor. Although such approaches certainly represent nanotechnology in action, their impact pales in comparison to how nanotechnology could potentially make a dream of certain environmentalists come true, namely, the "Hydrogen Economy." Hydrogen fuel cells use energy liberated when oxygen (from the air) and hydrogen combine to produce electricity. These fuel cells come in different sizes that could be used in cars, homes, offices, or mobile computing and communications devices. In theory, the hydrogen economy would have many advantages over the fossil fuel economy.

3.4 Nano-Solar Power

Like fuel cells, solar power has been on the verge of solving our energy problems for a long time, but somehow has never managed to actually do so. In part this is because both hydrogen and solar power share the fact that they are based on inexpensive, but low energy density fuels, so by the time they are delivered to the customer they are quite expensive. The promise of nanotechnology in both cases is that it will make both solar and hydrogen power much cheaper to do useful work.

There are actually (at least) three ways in which researchers have proposed harnessing the power of the sun: 1. ***Passive solar***. 2. ***Solar power stations***. 3. ***Photovoltaic (PV) systems***

3.5 The Nano-Enhanced Distributed Electricity Grid of the Future

The economics of any networked distribution system is a constant trade-off between using numerous hubs linked by short distances and a network in which a few large and powerful hubs linked by large distances. This trade-off is made by balancing the cost of transmission against the cost of the hubs. Much of the history of the telecommunications industry in the past 40 years could be written in terms of the shifting architectures as new technologies moved the balance from a few big switches to many small switches and back again.

4. NANOTECH, MEDICINE, AND THE PHARMACEUTICAL INDUSTRY

Nano-enabled solutions seem like a way to enable their goals through new types of regenerative medicine and better drug delivery. The pharmaceutical industry meanwhile is seeing a basic challenge to their core business model, which is based in

large measure on the concept of the "blockbuster drug," and these are apparently harder to find than they once were. Nanotechnology promises better drug discovery methods which could lead to new blockbuster drugs and through reinventing older blockbuster drugs with new delivery methodologies. NanoMarkets' research indicates that at the commercial level, product managers in the relevant departments at pharmaceutical firms are fairly skeptical about what nanotechnology can bring to the table that will help them improve the bottom line for their companies. This has been confirmed by other market research firms and only the biggest nano-boosters, mainly either executives at the start-ups or science writer types, really seem to believe that there will be huge early markets in nano-enabled medical products.

4.1 Seven Ways NanoEngineering Will Contribute to Nano-Enabled Drug Discovery

One area where nanotechnology seems most likely to have a short-term impact is in drug discovery. NanoMarkets research indicates that the pharmaceutical industry faces increasingly challenging market conditions that are leading to an intensified search for better drug discovery technologies. Nanotechnology can help with that search. The pharmaceutical industry has to discover and develop innovative medicines for a wide range of diseases in a marketplace that is both likely to experience growing regulatory and pricing pressures and that is increasingly targeted towards diseases that have been traditionally resistant to pharmacotherapy (e.g., solid tumors).

1. Atomic Force Microscopy, 2. Near-Field Scanning Optical Microscopy (NSOM), 3. Surface Plasmon Resonance, 4. Nano Mass Spectrometry, 5. Dip-Pen Nanolithography, 6. Nanoarrays, 7. Quantum Dots

4.2 Nanotechnology and Regenerative Medicine

The aging of the population in the Western world and in Japan would be a key driver for new medical technologies and particularly for nanomedicine. This is most obviously the case in the area of regenerative medicine, the restoring or replacing damaged tissues, bones, and organs. Some of this regenerative medicine is already available, while others are coming soon while some might be futuristic. Some of it is the nano-equivalent of Cosmetic surgery. Some of it is miracle medicine for the severely injured and disabled.

1) Nanogels: One of the applications in regenerative medicine for nanomaterials are gels that provide structures much like trellises used to "train" plants over which damaged nerve (and other) cells can grow as they regenerate. Similar gels have been on the market for a while, which should help the nanogels find acceptability quickly. The advantage that nanoengineering brings to the table here is that a nanoscale trellis fine-tunes the regenerative process, so that much of the original functionality is regained.

2) Organ Replacements: Completely artificial organs have been around for several decades, but have never really become as popular as once thought. The biggest example of this is the artificial heart. Meanwhile, there is a long wait for transplants of human organs and the procedures for making these transplants are lengthy, expensive, and dangerous. It is possible that new nanomaterials, coupled with nanoelectronic devices could make very significant contributions to constructing artificial organs that would go some way at least to make them a practical alternative to human transplants and thereby alleviate the shortage of organ donors. A more interesting, and perhaps, even a more likely direction, is to use "nanotrellises" of the kind described above to grow complete organs. According to one report, NASA has used this approach to grow heart cells and connect them up in a way that actually allows them to "beat" when put in the correct artificial environment.

3) Better Blood: A number of other less dramatic nano-enabled procedures should also help improve the cardiovascular system. There has been talk of creating artificial blood cells, which would consist of nanospheres filled with high-pressure oxygen that could be injected into the body. This would be as much as a drug delivery system and could be used to help treat heart attack or stroke victims. Or it could be used to enhance performance, perhaps becoming the next big business opportunities after oxygen bars? Incidentally, artificial blood cells already have a name: respirocytes.

4) Improved Memory: Primarily aimed at Alzheimer's patients, there are now several nano-related efforts to improve memory. These could also ultimately find a way into improving memory for the general population, if they proved to be effective and not too expensive. Much like the heart example above the choice is between a completely artificial solution using some kind of computer memory (nanomemory or not) and a nanoengineered conductive polymer to make the connection to the brain. Another possibility would be to take real neurons and plant them in the brain using some kind of nanoengineered vehicle based (perhaps) on some of the drug delivery systems that we have already met. In this particular case, the nanopackaging for the neuron would have to be such that the neuron could be artificially stimulated by a source (perhaps a nanobattery) in order to enable it to emit chemical neurotransmitters.

5. ENHANCING SECURITY WITH NANOTECHNOLOGY

Nanotechnology is not only useful for physical Technical applications for safety but also to help researchers use physics theories and to show application security strength. In a sense, this is similar to public key cryptography which uses theories of mathematical construct cryptographic applications and prove their level of encryption. Nanotechnology also helps to solve the challenge of devising a means of principles and run its precision by assessing the limits of the adversary. The idea of enlisting physical theories applications and security design prove their strength is not new. For quantum cryptography, for example, Information Technology Security facilitate secure exchange of secret keys [7].

5.1 A revision of two fundamental applications tamper-proofing and functionality obfuscation reflects nanotechnology's ability to deliver secure systems and scientifically establish their security robustness.

5.1.1 Tamper-proofing

A secured system is as strong as its weakest link. Key protection should include a method to prevent adversaries from reading the memory that stores the key from the outside. In addition, adversaries shouldn't be able to read the data bus that carries the key from storage to the processing element during the circuit's dynamic operation. An additional threat concerns probing the secret key at the processor's initial logic gates. Technical considerations that are more art than science,

however, drive current approaches to tamper-proofing for example, special coating techniques and clock irregularities. Instead of providing fundamental proofs, these approaches rely on the designer's instincts.

5.1.2 Functionality obfuscation

The possibility of hiding a program or logic circuit's functionality carries major security implications. The hidden functionality is regarded as a secret key known only to the designer. When trying to obfuscate a logic circuit's functioning, adversaries can "shave" a circuit layer by layer and copy the fabrication masks. They can then reconstruct the circuit's logic gates and analyze its functionality. Analysis leads to the conclusion that the discrete structure of programs and logic circuits inherently prevents true obfuscation. Yet, security officials constantly conduct practical trials intended to provide functionality obfuscation [8].

5.2 Nanotechnology and Tamper-Proofing

Tamper-proofing should be based on a scientific approach in which probing destroys the value the probe attempts to read. The energy the probe radiates or absorbs during the reading attempt should disrupt the mechanism that stores, conducts, or processes a bit, based on established physics laws. Nanotechnology might provide the right tool because the miniaturization gets to a level of handling a few electrons, increasing the probability of disturbing the bit's presentation merely by attempting to probe it.

5.2.1 Storage

Nonvolatile molecular memory uses a molecule to store a charge. Appropriate electrical bias is applied to set and reset a memory cell. In multilevel molecular memory, the electrical conductance in nanowires is adjusted by molecules that accept or emit electrons. Such principles provide ample possibilities for preventing external probing, whereby interaction between the probe and the molecule disrupts the stored charge or the electrons that control the nanowire conductance[10]. An external tampering probe radiates or absorbs energy needed for reading a stored value. Researchers can use rules from physics to exactly evaluate the meaning of the interaction between the probe and the charge or electrons that support the memory. Using formulas from physics for security analysis is comparable to using algebraic formulas to establish the complexity of breaking a mathematical code.

5.2.2 Data-bus tampering

Data-bus tampering involves techniques such as using an ion beam to drill a hole to the signal line, which the adversary then fills with a conducting material to bring the signal to the surface, where it's probed. The following three possible nanotechnology considerations pertain to preventing databus readings.

5.2.3 Quantum-dot cellular automata

A quantum-dot cellular automata (QCA) cell consists of four quantum dots positioned at the vertices of a square. Two added electrons occupy the cell's diagonals. Two possible polarizations can encode a bit. The interaction forces neighboring cells to synchronize their polarization and settle at a minimum energy-stable state. This forms the basis for conduction line and logic gates. Consider a case in which the system transfers a stored secret key from storage to the processor over a QCA conduction line. Probing the conduction line from the outside might cause instability in the states of the cells that are based on single-electron interactions[10].

5.3 Nanotechnology and Functionality Obfuscation

Fundamental studies argue that it's inherently possible to obfuscate software code or logic hardware functionality due to the discrete form of such structures. Based on fundamental observations from physics, these studies also point out that nanotechnology may form the foundations of hardware obfuscation, in which the circuit's discrete structure is blurred.

QCA

In standard microelectronics, the transistors and conduction lines are made of different substances, and different masks introduce them into the process. Since reverse engineering can recover the masks, it can also recover the logic gates and circuit functionality.

QCA cells are the building blocks of logic gates as well as the conduction lines that join the gates when forming a logic circuit. Making the gates and conduction lines of the same cells obfuscates the discrete logic structure. When observing the production masks, an adversary can't know whether a specific point in the circuit is part of a gate or part of the conduction line that joins the gates.

Chemical or biological processing

Leading nanotechnology research focuses on the fabrication of logic gates based on chemical or biological processes. Based on proved theories from chemistry and biology, we can expect to use a combined logic process that obfuscates internal functionality. A possible leading component here is the Fredkin gate, a universal logic gate that has great potential for being constructed by chemical or biological means. It also provides for reversibility, an attractive cryptographic feature. It's possible to use nanotechnology principles in security applications, based on physics practice and theories. The practice helps build the applications, while the associated theories provide a formal means to analyze security robustness. Research activities in these and similar directions might yield fundamental results.

6. FUTURE DEVELOPMENTS OF NANOTECHNOLOGY

Upstream engagement appears to encourage the public to focus on imagining positive outcomes for nanotechnologies. We raise issues in relation to the purpose of new science, such as nanotechnology. This section puts forward a model where human need rather than just consumer wants might influence the development of nanotechnologies. The chapter ends with some speculation about future directions that are desirable if the social and ethical concerns of society are to be met.

6.1 Nanotechnologies and Developing Countries

We live in a rapidly changing world. Technological advances are increasing productivity and income, quality of life, and life expectancy in the developed world, that is. The truth is that technological development is focused on meeting the wants of rich consumers. Scant attention is paid to the vital *needs* of people in the developing world. Each new technology that comes along tends to result in a wider gap between the rich and the poor in the world. Yet some innovations fail to be applied in developing countries where there is the need. the challenge is to

ensure that nanotechnologies are applied to areas of need in developing countries.

With existing technologies this becomes a challenge because the business models, including the supply chain logistics are already well established. In the case of new technologies there is a window of opportunity before products are released into the market to negotiate new business models.

In a global economy, many topical issues - for example, sustainable development, climate change, and democracy - are all influenced by the role of science and technology in society. A major challenge is to release public value from science and technology and to channel that public value into developing countries to help reduce poverty [3]. The challenge faced might be reframed as being one of "how do we enable nanotechnologies to deliver products which fullfil human needs rather than consumer wants?"

6.2 How Can Nanotechnologies Deliver Public Value?

The role of technology in development is perhaps even more important in the new century than it was in the last. In the era of globalization, *new technologies* are rapidly reshaping the livelihoods and lifestyles of people throughout the world. The pace of technological change is increasing, and is beyond the capacity of society to understand and regulate its impacts when the implications are profound and far reaching, as is the case with nanotechnologies. Most scientific and technological research is now in the private sector, producing research for Northern wants rather than Southern needs. Small-scale farmers and the informal sector give little attention to small-scale technological innovation.

Knowledge and communication-based industries are rapidly reshaping the global economy. Many believe that these trends are contributing to a new "knowledge divide" between the information-rich and the information-poor. There is an increasing sense of urgency-in the North and in the South-over the need to regain control over the ways *nanotechnologies* are developed and used. It is not recognized widely enough that the poor are able to innovate themselves, and innovations arising from developing countries need to be increasingly recognized and supported.

Traditional views of technology that rely on a linear model of innovation and diffusion are not appropriate to programmers that aim to respond to new technologies**.** The predominant traditional view has been based on technological determinism. As suggested, "the adoption of a particular technical system requires the creation of a particular set of social conditions as the operating environment of that system."

Yet we know that the technology for providing clean water has been known about and in use for thousands of years (e.g., the Romans around 300 BC). Failure to solve the issue might also be seen as a cultural or indeed political or managerial problem.

6.3 Balancing Risk and Opportunity

The convergence of the newly emerging technologies of the twenty-first century has the potential to revolutionize social and economic development and may offer innovative and viable solutions for the most pressing problems of the world community and its habitat. However, a better understanding of the potential benefits and hazards of nano-scale science and technology is essential. Materials fabricated on the nanoscale have properties that are different from those that are manufactured at a normal scale. For example, the precise way in which the atoms are arranged often leads to unusual optical and electrical properties. Carbon at the nanoscale can conduct electricity better than copper. In other cases the small size may have the effect of being more toxic than normal. A distinction can be made, in terms of risk assessment, between active and passive nanoparticles. Passive particles, such as a coating, are likely to present no more or less a risk than other manufacturing processes according to French. However, she goes on to assert that in the case of active nanoparticles, their ability to move around the environment leads to risks associated with control and containment.

7. CONCLUSIONS

In this paper, we have presented a detailed study of nanotechnology and its applications along with its advantages and disadvantages.

8. REFERENCES

[1] Takeda Hall, "Nanophotonics**:** Beyond the Limit of Optical Technology", Uniersity of Tokyo, October 2004. www.nanonet.go.jp/english/exchange/file/report04pgab.pdf

[2] *Lawrence Gasman* "Nanotechnology applications and markets**",** Boston :Artech House, 2006.

[3] An article on "Atomic Force Microscopy" by Nano Science Instruments. http://www.nanoscience.com/education/AFM.html

[4] Linette M. Demers, " Nanolithography : Applications and Functional Extensions", July 2005. http://www.nanoscienceworks.org/publications/dekker/120035889/

[5] Leo P.Kouwenhoven, Charles M. Marcus, Paul L.Mceuen, "Electron Trnasport in Quantum Dots", 1997. http://marcuslab.harvard.edu/papers/KouwenhovenReview.pdf

[6] Alexandra, *"Nano,-Six Nanotechnology center Funded by the National Science Foundation"* , MIT Technology Review, December 2001.

[7] Ellenbogen Toller, "Nanotechnology : Scientific Issues & International perspectives"

[8] Dr James, "Nanoelectronics: International activity and National Security Perspectives."

[9] Benjamin Arazi, Ben Gurion ,"Enhancing security with Nanotechnology," IEEE Computer Society, Oct 2006. http://www..*ieeecomputersociety.org/10.1109/MC.**2006**.338.*

[10] G. L. Snider, A. O. Orlov, I. Amlani, X. Zuo, G. H. Bernstein, C. S. Lent, J. L. Merz,and W. Porod ,"Quantum-dot cellular automata: Review and recent experiments", Journal of applied physics, June 2006. http://www.*ieeexplore.ieee.org/iel5/4915369/5023107/05023160.pdf?arnumber.*

Selection of Servomotors and Reducer Units for a 2 DoF PKM

Hermes GIBERTI, Simone CINQUEMANI

Mechanical Engineering Department, Politecnico di Milano,
Campus Bovisa Sud, via La Masa 34, 20156, Milano, Italy

ABSTRACT

The paper describes the problem of selecting the motor and reducer units for 2 degrees of freedom parallel kinematic machine. Parallel kinematic machines attract researchers and companies, because they are claimed to offer several advantages over their serial counterparts, like high structural rigidity and high dynamic performance. The need to increase production capacity, while maintaining the quality standards, required the implementation of automatic machines performance ever higher. In this context, it is of strategic importance in the machine design phase the correct selection of the motor-reducer unit. Unfortunately, the choice of the electric motor required to handle a dynamic load, is closely related to the transmission choice. The selection of suitable motors and transmissions is carried on introducing some parameters that describe the performance of the motor, the power required by the system and the influence of the transmission mechanical characteristics on the machine performance.

1. INTRODUCTION

The evolution of electronics in recent years has led to a wide diffusion of electric drives and their control systems. The ready availability and low cost of electronic devices has allowed rapid diffusion of mechatronic applications, highlighting the need for appropriate methods for selecting a motor-reducer unit. These procedures must be at the same time accurate and easy to use, they should be able to identify the available alternatives, compare them and help the designer in choosing the most appropriate one for his needs.

The choice of the electric motor required to handle a dynamic load is closely related to the choice of transmission. This operation, in fact, is bound by the limitations imposed by the motor's working range and is subjected to several constraints that depend indirectly on the motor (through its inertia J_M) and on the reducer (through its transmission ratio τ), whose selection is the subject of this paper.

A methodology for choosing the gear motor in order to ensure maximum acceleration of the system and reduce execution time for a particular law of motion is presented in [1]. This article introduces the so-called problem of *inertia matching*, showing how best performance can be reached when the inertia of the load, referred to the motor shaft, coincides with the inertia of the motor itself.

In [2] a procedure for the selection of an AC synchronous motor with permanent magnets and its reducer, for a generic load, is shown. The authors use normalized torques, velocities, and transmission ratios to separate the load from the motor characteristics. By virtue of this normalization, the simulations for one standard motor ($J_M = 1$ [kg m^2]) are applicable to other motors. In [3], the same procedure is extended to all types of servomotors. This methodology produces a chart representing all the usable motors and their corresponding normalized transmission ratios range, but not the available and actually usable commercial transmissions.

In [5] the choice of the motor-reducer unit is analyzed with regard to the dependence on the law of motion used to operate a generic load, while [6] evaluates the gain in motor torque as a consequence of the optimization of trajectories and highlights the effect that a variable transmission ratio has on the performance of the machine.

An in-depth discussion on the problem of motor reducer coupling can be found in [7]. Although this work is very accurate, it is extremely hard to use in a real industrial situation. On the other hand, [8] proposes a simpler approach, that consists in creating a database including the commercial motors and reducers and then trying all possible combinations.

In [9], [10] the choice of motor and reducer is made by comparing two parameters respectively related to motor features and to load demands. The relationships between motor and transmission are investigated by introducing some easily calculated factors useful for comparing all the available motor-reducer couplings and selecting the best solution. The procedure is carried on with the use of graphs that allow allow showing all the possible alternatives. Following this approach the paper shows how it is possible to select the motor-reducer units for a 2 d.o.f. parallel kinematic manipulator.

The paper is structured as follows. Section 2 gives a brief description of the manipulator with its main subsystems. Section 3 recalls the conditions to select a servomotor and a speed reducer and the corresponding checkouts. Section 4 introduces the multibody model of the system, developed to calculate the inverse dynamic. Section 5 shows the selection of the motor reducer units and the corresponding checkouts. Finally conclusions are drawn

in Section 6.

2. THE SYSTEM

The parallel kinematic machine under study is a 5R-2 d.o.f. manipulator consisting on 4 links (5 considering the ground) connected by five revolutionary joints (R) two of which are located on the ground and driven by motors (Fig.1). The manipulator can reach a place inside its workspace of coordinates $x = [x_e; y_e]^T$ which is as function of the actuated joints coordinates $q = [\theta_1; \theta_2]^T$.

It is constituted by 4 main elements:

1. the support, which is fixed and connected to the ground;
2. the driving system, constituted by 2 brushless motors, each actuating a joint. Main features of each motor are resumed in Tab.1

Table 1: Motor main features

Symbol	Description
T_M	motor torque
J_M	motor moment of inertia
$T_{M,N}$	motor nominal torque
$T^{TH}_{M,max}$	motor theoretical maximum torque
$T_{M,max}$	servo-motor maximum torque
ω_M	motor angular speed
$\omega_{M,max}$	maximum speed achievable by the motor
$\dot{\omega}_M$	motor angular acceleration
α	accelerating factor [9]

3. the transmission, which changes the torque and the speed supplied by the motor to the ones requested at joints. Main features of transmission are resumed in Tab.2

Table 2: Transmission main features

Symbol	Description
τ	transmission ratio
η	transmission mechanical efficiency
J_T	transmission inertia

4. the manipulator (coloured with light blue), machine consisting in 4 connected by five revolutionary joints

The task of the robot is defined by the user and generally can consists on:

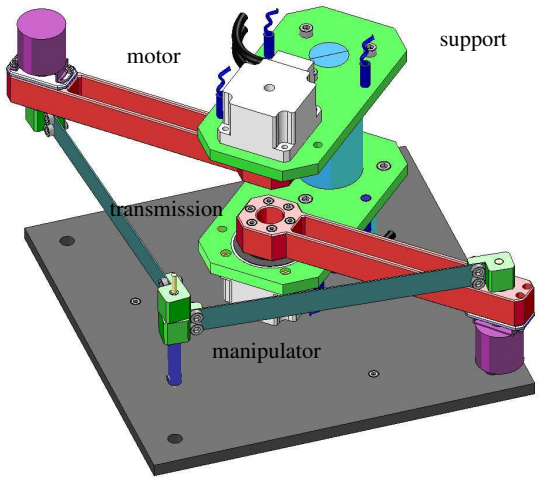


Figure 1: The manipulator

- pick and place operation;
- execution of a job along a defined trajectory.

The power supplied by the motor depends on the external load applied (T_L) and on the inertia acting on the system ($J_L\dot{\omega}_L$). Since different patterns of speed (ω_L) and acceleration ($\dot{\omega}_L$) generate different loads, the choice of a proper law of motion is the first project parameter that should be taken into account when sizing the motor-reducer unit.

Otherwise, it may be that the law of motion has been already defined and therefore represents a problem datum, and not a project variable. Once the law of motion is defined, all the characteristics of the load are known (Tab.3).

3. CONDITIONS OF SELECTION AND CHECKOUTS

As it is well known [2], the selection of the motor-reducer unit means checking the following conditions:

- rated motor torque:

$$T_{M,rms} = \sqrt{\frac{1}{t_a}\int_0^{t_a} T_M^2 \mathrm{d}t} \leq T_{M,N}; \qquad (1)$$

- maximum motor speed:

$$\omega_M \leq \omega_{M,max}; \qquad (2)$$

- maximum servo-motor torque:

$$T_M(\omega_M) \leq T_{M,max}(\omega_M)\,. \qquad (3)$$

Table 3: Load main features

Symbol	Description
T_L	load torque
J_L	load moment of inertia
T_L^*	generalized load torque ($T_L^* = T_L + J_L\dot{\omega}_L$)
$T_{L,rms}^*$	generalized load root mean square torque
$T_{L,max}$	load maximum torque
ω_L	load angular speed
$\dot{\omega}_L$	load angular acceleration
$\dot{\omega}_{L,rms}$	load root mean square acceleration
$\omega_{L,max}$	maximum speed achieved by the load
t_a	cycle time
β	load factor [9]

Once the available motor-reducer units have been identified it is necessary to check some further conditions:

- The maximum torque supplied by the servo-motor for each angular velocity achieved:

$$T_{M,max}(\omega_M) \geq \max\left|\frac{\tau T_L}{\eta} + \left(\frac{J_M + J_T}{\tau} + \frac{J_L\tau}{\eta}\right)\dot{\omega}_L\right| \quad \forall\omega. \tag{4}$$

- the effect of the transmission mechanical efficiency (η) and its moment of inertia (J_T) on the root mean square torque:

$$T_{M,N}^2 \geq T_{M,rms}^2 = \int_0^{t_a}\frac{T_M^2}{t_a}\mathrm{d}t = \\ = \int_0^{t_a}\frac{1}{t_a}\left(\left(J_M + J_T\right)\frac{\dot{\omega}_L}{\tau} + \frac{\tau T_L^*}{\eta}\right)^2\mathrm{d}t \tag{5}$$

- the resistance of the transmission as supplied by the manufacturer.

Conditions expressed by inequalities (1), (2) and (3) are well known in the literature and represent the starting point of all the procedures for motor and reducer selection.

This paper follows the procedure of selection described in [9], [10] where these conditions have been rewritten by introducing certain parameters related to motor and load features. Most important are the accelerating factor α:

$$\alpha = \frac{T_{M,N}^2}{J_M}, \tag{6}$$

and the load factor β:

$$\beta = 2\left[\dot{\omega}_{L,rms}T_{L,rms}^* + (\dot{\omega}_L T_L^*)_{mean}\right] \tag{7}$$

It is important to highlight that all the parameters used have a physical meaning and are easily obtained. In this way the designer will have a clear idea of the needs and of the steps to follow.

The preliminary choice of motor is made by comparing only the values α and β; these values are easily calculated if we know the mechanical properties of the motor and the load features. A motor must be rejected if $\alpha < \beta$, while if $\alpha \geq \beta$ the motor can have enough rated torque if τ is chosen properly.

Once these parameters are calculated, conditions (1), (2), can be easily expressed, for each considered motor, as functions of τ.

$$\tau_{min}, \tau_{max} = \\ = \frac{\sqrt{J_M}}{2T_{L,rms}^*}\left[\sqrt{\alpha - \beta + 4\dot{\omega}_{L,rms}T_{L,rms}^*} \pm \sqrt{\alpha - \beta}\right]. \tag{8}$$

$$\tau_{M,lim} = \frac{\omega_{L,max}}{\omega_{M,max}} \tag{9}$$

The result is a range of acceptable transmission ratios for each motor.

$$\tau_{max} \geq \tau \geq \max\left(\tau_{min}; \tau_{M,lim}\right). \tag{10}$$

However, since it is difficult to express the constraint imposed on the servo-motor maximum torque (3) as a function of τ, this condition will be checked after the motor and its transmission have been chosen.

4. MULTIBODY MODEL

To evaluate the power required to the motors when the manipulator performs the desired task, a multibody model has been developed.The model allows to define the trajectory the robot has to follow, the end effector motion law and the external forces applied to the manipulator.

Once the task has been defined, simulations allow to calculate:

- torques required to the motor-reducer unit as a function of time ($T_M(t)$). From this result it is possible to calculate the root mean square torques ($T_{M,rms}$) and the maximum torque ($T_{M,max}$) required during the task;
- the motor angular velocity as a function of time ($\omega_M(t)$);
- reaction forces on joints to properly size the structure.

For example it is possible to define a linear trajectory (Fig.2), with $trapezoidal$ motion law ($\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$) (Fig.3), with a cycle time $t_a = 0.8s$.

The corresponding motion laws of joints and the required motor torques are depicted in Fig.4,5.

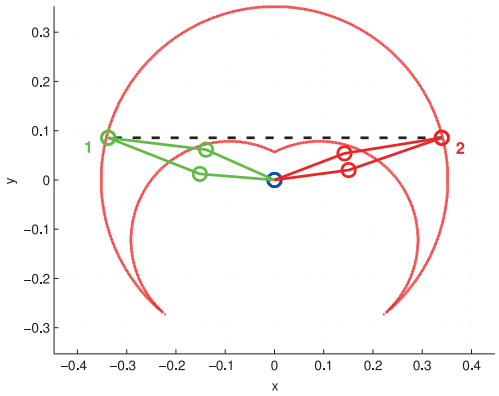


Figure 2: Linear trajectory (diametrical)

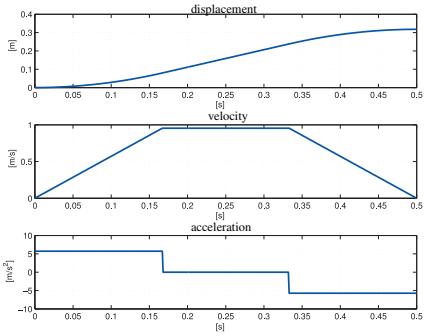


Figure 3: Motion law along the desired trajectory

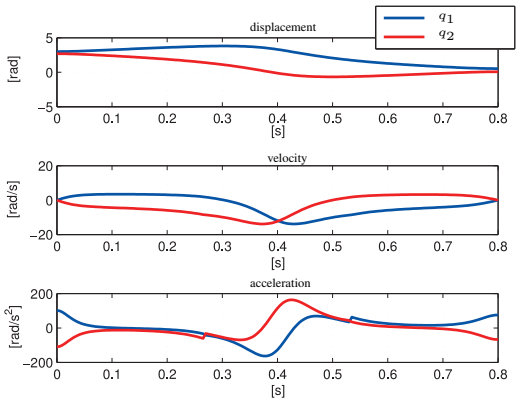


Figure 4: Motion law $q_1(t)$, $q_2(t)$

5. MOTOR REDUCER UNIT SELECTION

The choice of the motor reducer unit should satisfy some design requirements:

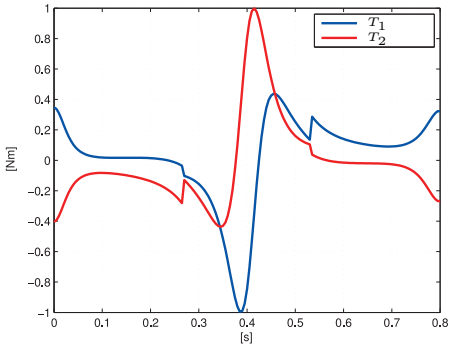


Figure 5: Required motor torques

- limited space and weight;
- high positioning accuracy and repeatability;
- high torque availability;
- low wear;
- high torsional stiffness;
- high transmission ratio.

For these reasons DC brushless motors and Hermonic Drive speed reducers have been considered. Available motors are models: $FHA-8C$, $FHA-11C$, $FHA-14C$, while speed reducers have transmisison ratios: $1/30$, $1/50$, $1/100$.

Once the task the robot has to perform (trajectory, mnotion law, external forces applied), numerical simulatuions on multibody model allows to calculate all the terms related to the load as, generalized resistant toque (T_L^*) applied to motor shafts, velocities ($\dot{(q)}_1$,$\dot{(q)}_2$) and acceletions ($\ddot{(q)}_1$,$\ddot{(q)}_2$). using these values it is possible to calculate the load factor β that can be compared with the accelerating factors α calculated for each considered motor using information on manufacturers catalogues.

Consider now the motor task described in Section 4. Figure 6 shows the values of α and β, highlighting that all the considered motors can be profitably used to perform the required task since the condition $\alpha \geq \beta$ is verified.

For each motor it is possible to calculate the range of suitable transmission ratios using eq.(10). Figure 7 highlights that transmission ratios $1/50$ e $1/100$ can not be selected becaouse of the limit on the maximum achievable speed. On the other side it results that motor *FHA-11C*, coupled with the transmission ratio $1/30$ is a good combination to perform the task.

However, it may be simplistic to choose the motor only on a specific task, since the robot should be suitable for many movements. It is also desirable to optimize the choice of the motor-reducer unit, investigating which solution can

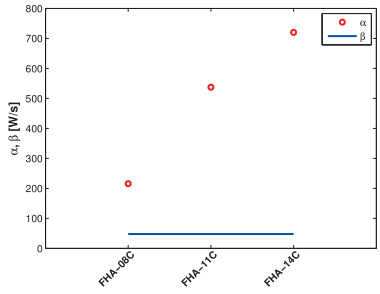


Figure 6: α and β

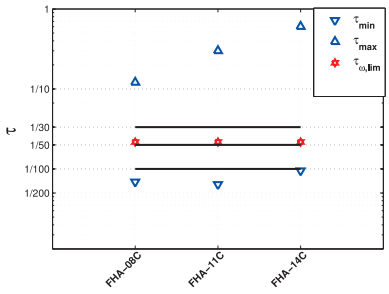


Figure 7: Ranges of suitable transmission ratios τ

offer the best performance, ie which allows smaller cycle times with the same law of motion.

Considering three simplified trajectories (radial, diametrical and circular movement) and maintaining a trapezoidal law of motion, one can gradually decrease the cycle time to obtain the limit transmission ratio.

Figure 8 shows the trend of the load factor β when the cycle time decreases. This value can be compared with the accelerating factors α calculated for each motor. The graph gives an overview to easily understand when a motor can be no longer suitable to perform the task along a certain trajectory. Figure 9 shows the maximum speed required by the task for each trajectory. It can be compared with the maximum speed achievable by each speed reducer allowing to identify which is the most suitable transmission ratio for the purposes.

Looking at Fig.8 and Fig.9 it is evident that, for every considered machine task, motor $FHA-11C$ and transmission ratio $1/30$ represent alway the best solution to perform the task. Once the motor reducer unit has been selected, checkout described in Sec.3 have to be performed. In particular, Figures from 10 to 12 shows the limits of the selected motor reducer unit in terms of maximum torque and nominal torque exertable. By reducing the cycle time

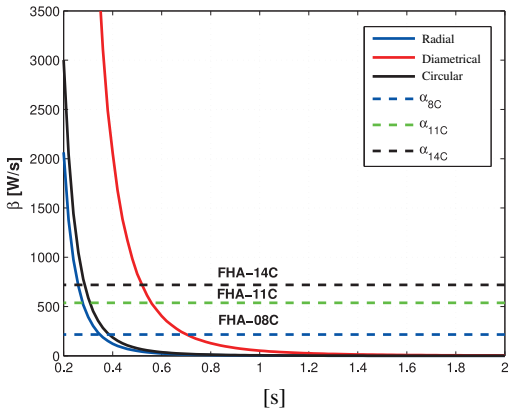


Figure 8: Comparison between α values and β as a function of the cycle time

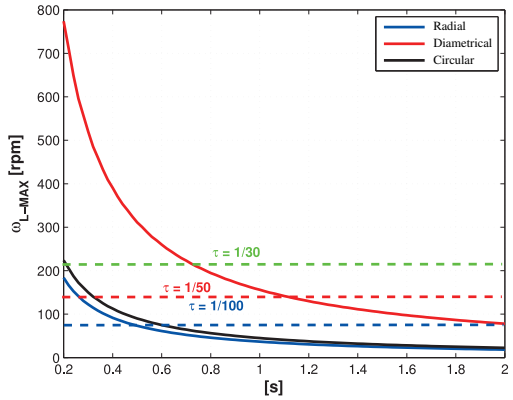


Figure 9: Comparison between ω_{L_max} values and τ as a function of the cycle time

the points (T_{rms},ω_{rms}) and (T_{max},ω_{max}), move to the border of the motor working zone. The closest point represent a limit working condition.

6. CONCLUSION

The paper presents a the selection of the most suitable motor and speed reducer pairing to drive a parallel kinematic machine, once its task has been defined. The methodology is based on information achievable on manufacturers' catalogues and it is carried out through a graphical representation of the characteristics of the machine and of available motors and speed reducers. The designer has then a useful procedure to compare all the feasible solutions and to choose the best one.

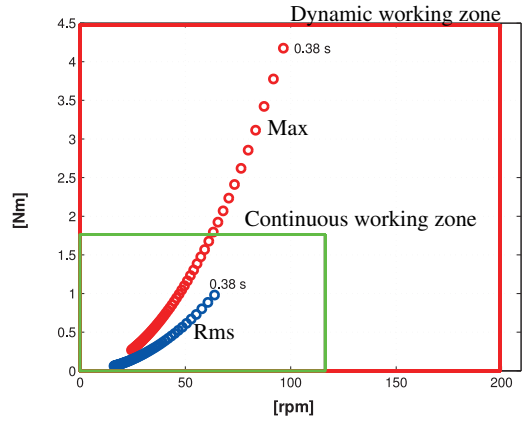


Figure 10: Checkout for a task with radial trajectory

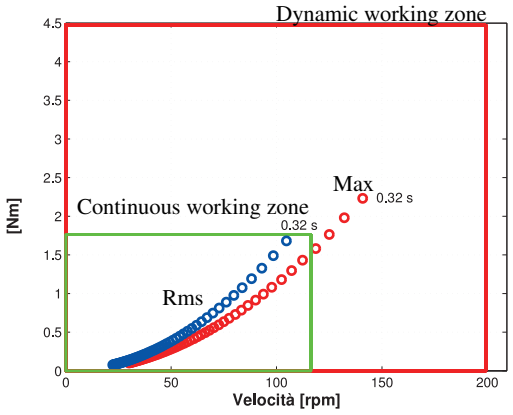


Figure 12: Checkout for a task with circular trajectory

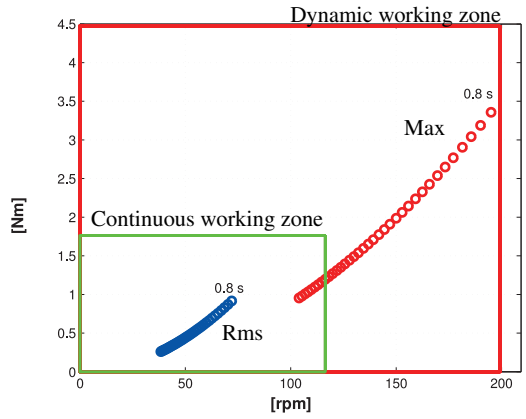


Figure 11: Checkout for a task with diametrical trajectory

REFERENCES

[1] Pasch K.A.,Seering W.P. (1984) On the drive systems for High-performance Machines, Transactions of ASME 106:102-108

[2] Van de Straete H.J, Degezelle P., de Shutter J., Belmans R.(1998) Servo Motor Selection Criterion for Mechatronic Application, IEEE/ASME Transaction on mechatronics 3:43-50

[3] Van de Straete H.J, de Shutter J., Belmans R. (1999) An Efficient Procedure for Checking Performance Limits in Servo Drive Selection and Optimization, IEEE/ASME Transaction on mechatronics 4:378-386

[4] Cusimano, G., "A procedure for a suitable selection of laws of motion and electric drive systems under inertial loads", *Mechanism and Machine Theory* 38, 519-533 (2003)

[5] Cusimano G. (2003) A procedure for a suitable selection of laws of motion and electric drive systems under inertial loads, Mechanism and Machine Theory 38:519-533

[6] Van de Straete, H.J, de Shutter, J., Leuven, K.U., "Optimal Variable Transmission Ratio and Trajectory for an Inertial Load With Respect to Servo Motor Size", *Transaction of the ASME* 121, 544-551 (1999)

[7] Cusimano G. (2007) Optimization of the choice of the system electric drive-device-transmission for mechatronic applications, Mechanism and Machine Theory 42:48-65

[8] Roos F., Johansson H., Wikander J. (2006) Optimal selection of motor and gearhead in mechatronic application, Mechatronics 16:63-72

[9] H. Giberti, S. Cinquemani, G. Legnani (2011) A Practical Approach for the Selection of the Motor-Reducer Unit in Electric Drive Systems, Mechanics Based Design of Structures and Machines (in press)

[10] H. Giberti, S. Cinquemani, G. Legnani (2010) Effects of the mechanical characteristics of the transmission on the choice of motor-reducers, Mechatronics 20(5):604-610

A Geometrical Index to Evaluate the Kinematical Properties of a 2 Dof Parallel Kinematic Manipulator

Hermes GIBERTI, Simone CINQUEMANI, Stefano AMBROSETTI

Mechanical Engineering Department, Politecnico di Milano,
Campus Bovisa Sud, via La Masa 1, 20156, Milano, Italy

ABSTRACT

High dynamic performances, stiffness, accuracy and precision are just some main features that make the Parallel Kinematic Machines (PKM) more suitable and desiderable than serial robots for some typicall industrial applications such as laser cutting, pick & place operations and machining. The development of PKM brings to a process of optimization meant to improve the kinematic behaviour of the machine. This paper presents a new practical index, based on isotropy condition, easy to use for the design and project of 2-dofs parallel manipulators.

keywords: parallel manipulator, kinematic optimization, isotropy, performance index, optimum design.

1. INTRODUCTION

Industrial reality is deeply interested in Parallel Kinematic Machines thanks to their closed loop structure that provides higher stiffness and lower moving masses -and therefore lower inertia effects- than their counterparts, the serial manipulators, allowing to reach great advantages, among which higher dynamic performances and better positioning accuracy. However these machines carry on a serious drawback which is related to their workspace and nonlinear dynamic equations. In fact, a PKM has usually a complex geometrical workspace shape and its performances may vary significantly for different points in the workspace and for different directions.

In order to satisfy industrial demands, the PKMs must be designed pursuing the best workspace properties such as a regular shape and homogeneous kinetostatic performances. The optimum kinematic design is therefore mainly concearned with the definition of geometric parameters - i.e. the dimensions and the orientation of its links - that maximize these performances.

The kinematic and workspace optimization is typically analyzed following numerical approaches based on the concept of isotropy and the definition of performance indices calculated from the jacobian matrix.

This paper presents a new kinematic index based on geometrical considerations that deeply simplifies the design of one of the most used PKMs, which is the five-bar 2DOFs manipulator[1].

This machine is widely used and appreciated in industrial field. The interest is confirmed by the scientific literature regarding, in particular, the kinematic optimization of this PKM ([1], [8], [6], [2], [4], [5], [7]).

2. 5R 2DOF PKM

The 5R 2DOF PKM has the classic configuration shown in Fig. 1. The manipulator is characterized by a symmetric structure, made of four links and a fixed frame connected by five revolute joints. In order to place the motors on the fixed base the joints attached to the ground are the ones driven.

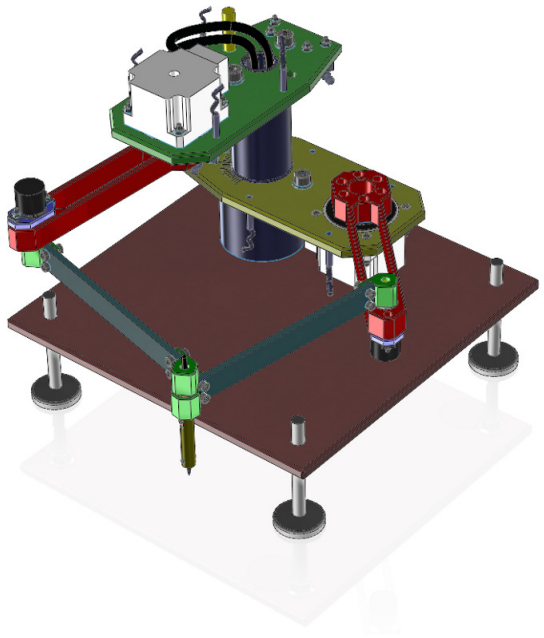


Figure 1: A typical 5R 2DOF PKM.

Looking at the scheme depicted in Fig.2 it is noted how, properly degenerating some joints, the manipulator

[1]Now called synthetically "5R 2DOF PKM"

can reach significantly different configurations even if similar from the kinematic point of view.

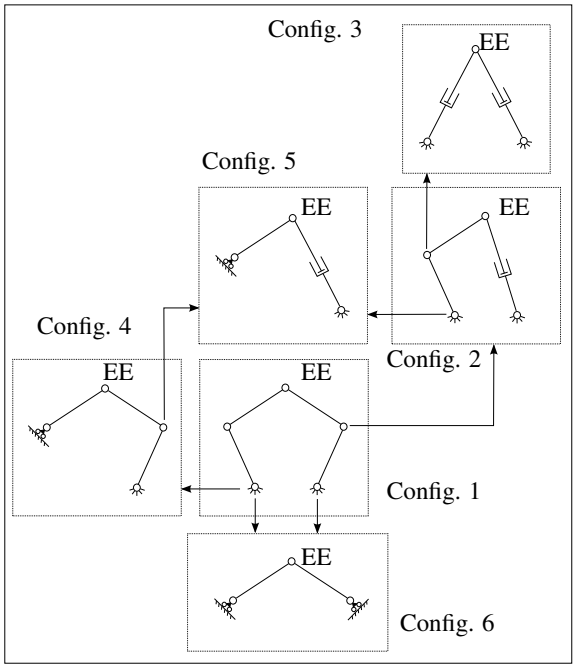


Figure 2: Different configurations of 5R 2DOF PKM.

This degeneration consists on making the axis of rotation of a revolute joint as improper, thus transforming it into a prismatic joint. The diagram depicted in Fig.2 can be read like this: each configuration can be achieved by making improper a joint; arrows indicate which axis of rotation is made improper and the corresponding configuration reached. The end effector of the manipulator is placed at one of the robot joints. It is assumed that this joint can not degenerate.

3. TYPICALL APPROACH

Isotropy

Isotropy is defined as the ability of the robot to have identical kinetostatic properties along each direction. The isotropic behaviour of a robot, as a function of its position inside the workspace, can be analyzed through some indices related to the jacobian matrix using the well known relationships:

$$\dot{Q} = J^{-1}\dot{S} \qquad F_s = J^{-T}F_a \qquad J = \frac{\partial S}{\partial Q} \tag{1}$$

where J is the jacobian matrix relating the gripper velocities $\dot{S}$ with those of the actuators $\dot{Q}$, as well as the general forces F_a exerted with the forces and the torques F_s applied to the gripper. Remembering that the i^{th} singular value $\sigma_i(J)$ of the jacobian matrix J is defined as the square root of the eigenvalues of the corresponding matrix $J^T J$:

$$\sigma_i(J) = \sqrt{\lambda_i(J^T J)} \qquad \text{with:} \qquad \lambda_i \geq 0 \tag{2}$$

The singular values $\sigma_i(J)$ of the jacobian matrix can be interpreted as Velocity Amplification Factors (VAF) between the actuated joints velocities and the end effector ones. The ratio between the maximum and the minimum singular value can be considered as an index of isotropy:

$$cond(J) = \frac{\sigma_{max}}{\sigma_{min}} = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \tag{3}$$

which is known to be the conditioning number of the jacobian matrix [11]. When it is verified $cond(J) = 1$, the minimum and the maximum eigenvalues coincide and VAF are equal. In this case the manipulator is defined as isotropic. Condition on isotropy can also be expressed as [10]:

$$J^T J = kI \tag{4}$$

where k is a scalar and I is the identity matrix. It means isotropy can be achieved when jacobian matrix is proportional to an orthogonal matrix.

Ellipses of manipulability

A similar result can be obtained considering the robot ellipses of manipulability for speed and force. They have been introduced to respectively approximate the locus of points of maximum speed or force achieved by the end effector of the robot [3]. They give a standard method, applicable also to complex manipulators with many degrees of freedom, to indicate along which directions the robot can reach high speed or exert high forces.

The evaluation of these ellipses for different reached positions shows the robot behaviour inside its workspace. The determination of the ellipses can be done on the basis of standard procedures that use the eigenvectors and singular values of the jacobian matrix (or matrices derived from it). Ellipses are defined according to the following assumption [3]: imagine the actuators have a total speed k_v to share between them with the constraint that the sum of the squares of the velocity is constant. From equation (1) one gets:

$$\left\|\dot{Q}\right\|^2 = k_v^2 = \dot{Q}^T\dot{Q} \tag{5}$$

$$\dot{S}^T J^{-T} J^{-1}\dot{S} \leq k_v^2 \quad \rightarrow \quad \dot{S}^T (JJ^T)^{-1}\dot{S} \leq k_v^2 \tag{6}$$

Once the maximum speed reached by each actuator is set, Eq. (6) defines an ellipse in the $\dot{x}\dot{y}$ plane described by the matrix $(JJ^T)^{-1}$. More precisely, the lengths of the principal axes of the ellipse, which correspond to the inverse of minimum and maximum eigenvalues of JJ^T, represent the minimum and maximum velocity amplification

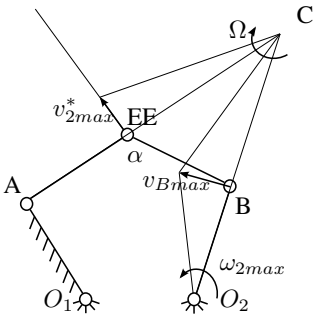


Figure 3: Maximum velocity along direction r

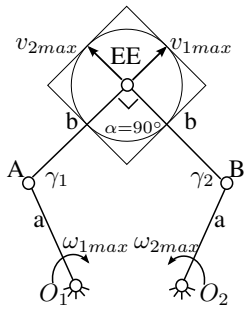


Figure 4: Ellipse of manipulability in velocity in isotropic configuration

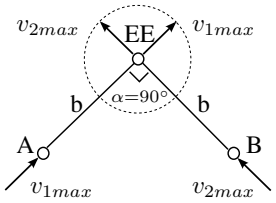


Figure 5: Graphical representation of conditions to reach isotropy

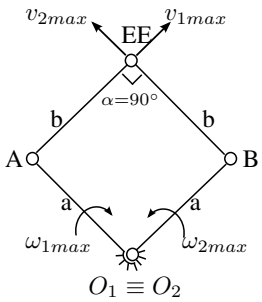


Figure 6: 5R2DOF manipulator with coaxial grounded joints

factor, while eigenvectors of $J^T J$ represent the direction of the principal axes of the ellipse. According to this representation, the ellipse of manipulability in velocity represents the locus of points of maximum velocity. Since generally one is interested in the shape of the ellipses rather than in its actual size, it is possible to assume $k_v = 1$. Due to kinetostatic duality [3], similar considerations can be done for ellipses of manipulability for forces. Remembering Eq. (4) the condition of isotropy can be interpreted by looking the ellipsoids of manipulability of the robot: according to this representation, isotropy is achieved when ellipses (or ellipsoids) become circles (or spheres). In this condition, in fact, the axes have the same length and then all the jacobian singular values are identical.

Global and Local Indices

The conditioning number and other indices like the determinant of jacobian matrix or the minimum singular value refer to an optimal condition (isotropy, manipulability or minimum stiffness) reached in a single point of workspace and they are therefore called *local indices*. On the other hand *global indices* have the purpose to describe the optimal condition of the whole workspace and they are usually calculated as a sort of average of local indices.

The simplest isotropy global index is in fact [11]:

$$k_G = \frac{\int_W k dW}{\int_W dW} \tag{7}$$

where W is the workspace and k is the local index, for example, the conditioning number.

The main goal of a global index is to supply to the weak information given by a local index, which allows to know the behaviour of a manipulator only in some poses. For this reason, global indices can be profitably used to design and optimize a robot since they are able to describe its behaviour whithin the whole workspace. Moreover they can be used to evaluate kinematic performances of existing manipulators.

4. A NEW INDEX

The conditions of isotropy for a manipulator can be graphically identified, considering the relative kinematic of the manipulator links. It has been said the ellipses of manipulability in velocity approximate the locus of points of maximum speed achievable by the end effector. To get that ellipse one can impose the EE to move along a given direction at its maximum speed by pushing up the motors. The contribution of each motor depends on the direction to follow. Consider the manipulator depicted in Fig. 3. Suppose, for example, one wants to move the end-effector of the manipulator along the direction r orthogonal to the link AC. To achieve maximum speed ($v^*_{2,max}$) in this direction the system will be forced to hold the motor 1, while moving the motor 2 at maximum speed ($\omega_{2,max}$). Knowing the pose of the robot one gets:

$$\underline{v}_{Bmax} = \underline{\omega}_{2max} \times \underline{O_2 B} = \underline{\Omega} \times \underline{CB} \tag{8}$$

where C is the instant center of the link B-EE and $\underline{\Omega}$ its absolute angular speed, whose modulus is:

$$|\underline{\Omega}| = \left| \omega_{2max} \frac{\underline{O_2 B}}{\underline{CB}} \right| \tag{9}$$

Finally the maximum speed along direction r is:

$$\underline{v}_{Bmax} = \underline{\Omega} \times \underline{EEC} \tag{10}$$

The reverse should be done if one wants to reach the maximum speed along a direction orthogonal to BC. By repeating this for all directions, the locus of points is represented by a polygon.

For convenience of representation, it is usual to approximate these polygons respectively with ellipses, or with circles (Fig.4). Ellipses are inscribed into the polygon and tangent to it in the middle point of each side. Obviously, depending on the configuration of the robot, the ellipse of manipulability in velocity has different shapes and sizes.

When the manipulator is in an isotropic configuration the polygon is a square and the corresponding ellipse degenerates into a circle. Looking at Fig. 4 and Fig. 5, the considered manipulator is in an isotropic configuration if, and only if:

$$\begin{cases} \alpha = 90^\circ & \rightarrow \quad \text{velocities are perpendicular} \\ \gamma_1 = \gamma_2 & \rightarrow \quad \text{velocities are equal} \end{cases} \tag{11}$$

These simple geometrical considerations can be extremely useful to evaluate the behaviour of the manipulator inside its workspace [1]. Once the condition expressed in Eq. (11) is verified, the triangle A-EE-B is defined. Its position and orientation to reach isotropy can be determined knowing the features of the actuators and the geometrical dimensions of the links in order to verify Eq. (11).

Consider now the standard configuration of the manipulator and assume that its links are symmetric. The maximum speeds achievable by the end effector, for a generic position, depends on the maximum speed exertable by the actuators. The conditions in Eq. (11) define the geometrical condition to assure isotropy (Fig. 5). These conditions are verified for a limited number of poses [13] that can not be reached without crossing a singularity configuration [12].

However, making the axes of rotation of the links connected to the ground as coaxial (Fig.6), the transmission angles are always equal and so are the speed v_1 and v_2.

In this configuration, the angle α becomes the only indicator of isotropic condition.

The manipulator depicted in Fig.6 is also the one that maximize the workspace area [12] and this is the reason to concentrate on this configuration.

In industrial applications, the solution with coaxial ground joints is feasible but requires particular attentions for the space of the motors; however, even if it is not a practicable way to collapse the ground joints, they are placed as clear as possible, in a configuration that tends to the ideal one.

In this configuration, it is possible to identify the so called isotropy loci represented by two concentric circumferences whose centers coincides with the grounded joints and whose radius R is:

$$R_{e,i} = b\frac{\sqrt{2}}{2} \pm \sqrt{a^2 - \left(b\frac{\sqrt{2}}{2}\right)^2} \tag{12}$$

As it is easy to notice, the isotropy is reached if $a > b\frac{\sqrt{2}}{2}$, otherwise the Eq. (12) is impossible. In particular,

1. if $0 < a < b\frac{\sqrt{2}}{2}$, there are no points of isotropy;
2. for $a = b\frac{\sqrt{2}}{2}$, two coincident isotropy loci exist. Their shapes are two circumferences with radius $R_{e,i} = b\frac{\sqrt{2}}{2}$;
3. if $b\frac{\sqrt{2}}{2} < a < b$, the mutual distance between the two circumferences increases for $a \rightarrow b$ (Fig.10);
4. if $a = b$, the circumferences have radius respectively $R_e = a\sqrt{2}$ and $R_i = 0$ (it means a isotropy locus degenerates into a point);
5. if $b \geq a$, the two circumferences where isotropy is reached still exist, but the manipulator must cross singular configurations to reach both the isotropy loci.

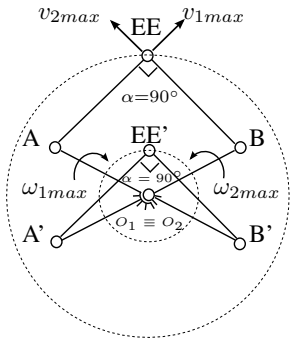


Figure 10: Continuum isotropy loci for PKM with coaxial grounded joints when $b\frac{\sqrt{2}}{2} < a < b$

It is clear that the most suitable condition is $b\frac{\sqrt{2}}{2} < a < b$. In order to know the optimized condition, the new index of isotropy $\delta = |90 - \alpha|$ is compared to a classical index ($\xi = 1/cond(J)$) based on the conditioning number of the jacobian matrix.

The comparison is shown in Fig. 7, where δ and ξ are depicted as a function of the position of the end-effector in radial direction, for different values of a/b.

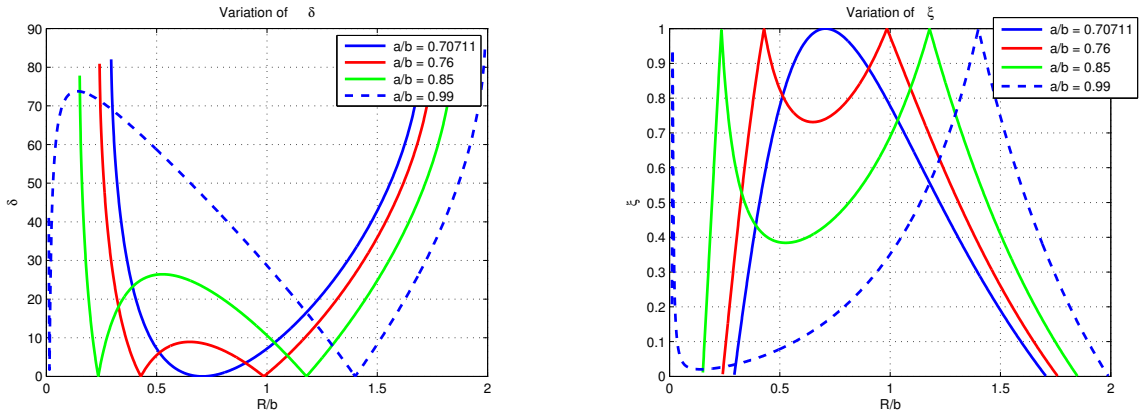
These two local indices are calculated for each position of the end-effector and averaged along the workspce for every value of a/b, as shown in Eq. (13) and depicted in Figures 8 and 9.

$$\begin{cases} \xi_{mean} = \dfrac{\sum_{x,y} \xi_{x,y}}{A_w} \\ \delta_{mean} = \dfrac{\sum_{x,y} \delta_{x,y}}{A_w} \end{cases} \tag{13}$$

Then global indices can be obtained as indicators of the mean behaviour of the manipulator.

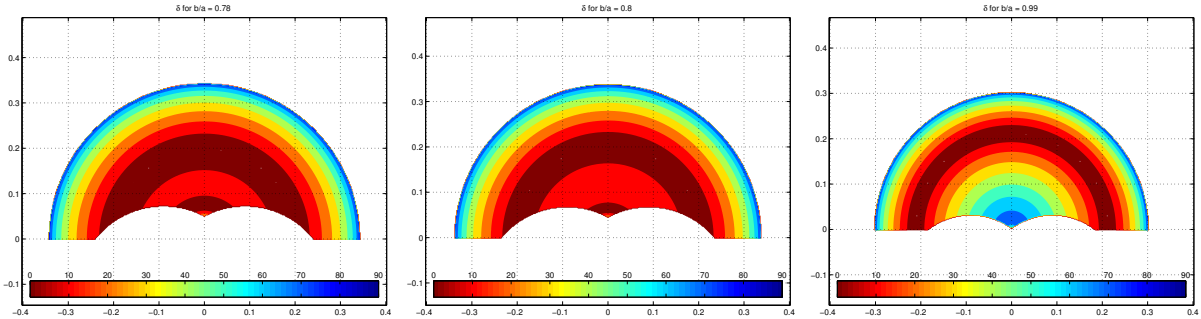
Thanks to the adimensional optimization, a general result is obtained and it can be used for the project of any 5R 2DOF PKM with coaxial joints connected to the ground.

The comparison brings to a similar conclusion:



(a) (b)

Figure 7: Variation of δ and ξ for different values of a/b along normalized radial direction.



(a) (b) (c)

Figure 8: Values of δ on the workspace for different a/b. a) $a/b = 0.78$; b) $a/b = 0.8$; c) $a/b = 0.99$.

- for ξ, $(b/a)_{optimum} = 0,759$;
- for δ, $(b/a)_{optimum} = 0,762$;

Due to their nature, the two indices δ and ξ give results slightly different: the angle α is a limited variable between two finite values while the $cond(J)$ may become infinite. Therefore the trend of these indices can have a different behaviour, even if both of them express the same condition.

5. CONCLUSIONS

In this paper a new alternative index representing the condition of isotropy of 5R 2DOF PKM with coaxial grounded joints has been presented. The information gained from this index has the same value of a classical index like the conditioning number. However, while the last one is hardly appreciable and has a difficult interpretation, the new index α allows to easily understand the beahiviour of the manipulator and its closeness to isotropy. This can be reached simply analyzing geometrical information.

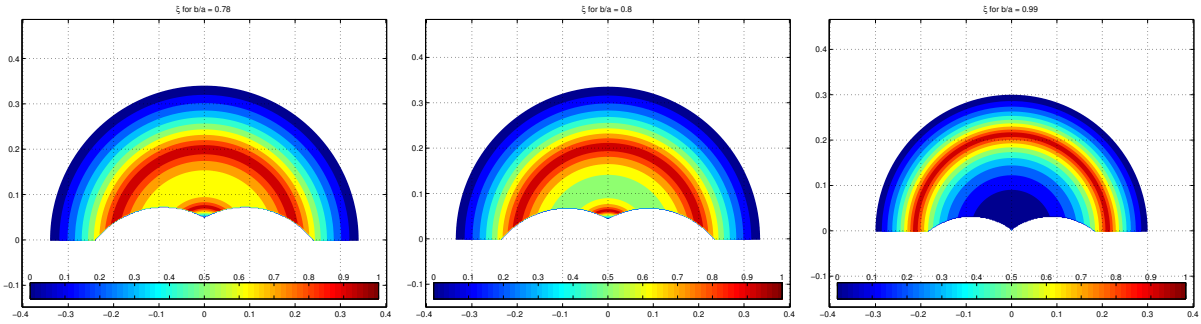
The index α can be used for a fast design and optimization of this kind of manipulators.

REFERENCES

[1] G. Legnani, D. Tosi, I. Fassi, H. Giberti, S. Cinquemani, " The point of isotropy and other properties of serial and parallel manipulators", *Mechanism and Machine Theory*, 45(10), (2010), 1407-1423.

[2] X.J. Liu, J. Wang, G. Pritschow, " On the optimal kinematic design of the PRRRP 2-DoF parallel mechanism", *Mechanism and Machine Theory*, 41 (2006) 11111130.

[3] T. Yoshikawa, "Foundations of Robotics: Analysis



(a) (b) (c)

Figure 9: Values of ξ on the workspace for different a/b. a) $a/b = 0.78$; b) $a/b = 0.8$; c) $a/b = 0.99$.

and Control", *MIT Press*, (1990) Cambridge, MA.

[4] X.J. Liu, J. Wang, "A new methodology for optimal kinematic design of parallel mechanisms", *Mechanism and Machine Theory*, 42 (2007) 12101224.

[5] X.J. Liu, J. Wang, H.J. Zheng, " Performance atlases and optimum design of planar 5R symmetrical parallel mechanisms", *Mechanism and Machine Theory*, 41 (2006) 119144.

[6] J. J. Cervantes-Snchez, J. C. Hernndez-Rodrguez, J. G. Rendn-Snchez, " On the kinematic design of the 5R planar, symmetric manipulator", *Mechanism and Machine Theory*, 36 (2001) 1301-1313.

[7] A. Fattah, A. M. Hasan Ghasemi, " Isotropic Design of Spatial Parallel Manipulators", *The International Journal of Robotics Research*, 21 (2002), 811.

[8] G. Alici, B. Shirinzadeh, " Optimum synthesis of planar parallel manipulators based on kinematic isotropy and force balancing", *Robotica* , 22 (2004), 97108.

[9] G. Legnani, " Robotica Industriale", CEA Ambrosiana Editor, 2003.

[10] J.P. Merlet, " Jacobian, Manipulability, Condition Number and Accuracy of Parallel Robots", *Journal of Mechanical Design*, 28 (2006), 199-206.

[11] C. Gosselin, " Stiffness Mapping for Parallel Manipulator", *IEEE Trans. on Robotics and Automation*, 6-3 (1990), 377-382.

[12] X. Liu, J. Wang, H. Zheng "Optimum design of the 5R symmetrical parallel manipulator with a surrounded and good-condition workspace ", *Robotics and Autonomous Systems*, 54 (2006), 221-233.

[13] K.Y. Tsai, P.Y. Lin, T.K. Lee, "4R spatial and 5R parallel manipulators that can reach maximum number of isotropic positions", *Mechanism and Machine Theory*, 43 (2008) 68-79.

Intelligent Engineering for Building Sustainability Beyond 2010

Bradley HEMPHILL
Electrical Engineering Solutions, bradley.hemphill@eeslive.com
Tokai, Cape Town, South Africa

and

Heidi BERGE
Electrical Engineering Solutions, heidi.berge@eeslive.com
Tokai, Cape Town, South Africa

and

Annabel EATON
Electrical Engineering Solutions, pr@eeslive.com
Tokai, Cape Town, South Africa

ABSTRACT

This paper serves as a case study on the economic and environmental sustainability of new buildings, in particular referring to five of the major Information and Communications Technology (ICT) and Building Automation System (BAS) projects either designed and/or project managed by Electrical Engineering Solutions (EES), for the Soccer World Cup 2010 hosted in South Africa.

To understand sustainability of new buildings, it is useful to know what the threats to sustainability are, and these are discussed in the context of the five projects below. Going forward, a "sustainability ideology" should be key for developers and other stakeholders. ICT and BAS professionals must engage with architects and construction consultants at an early stage to ensure that the required infrastructure facilities are incorporated in the architectural design.. Furthermore the ICT industry needs to educate and up-skill the built environment delivery sector on a continuous basis, at both student and professional levels.

In general, it can be shown that an intelligent building design team interacts to set energy performance goals, manage the building system's interdependencies and produce a more efficient, cost-effective, smarter building.

The implementation of a smart building design is usually best delivered by a master systems integrator (MSI).

Keywords: Sustainability, Building Automation Systems, Convergence, Information Technology.

1. INTRODUCTION

This case study illustrates how various convergence ICT principles (all ICT systems converged on one physical network) and engineering design principles aim to combat the challenges regarding sustainability in each of the following five projects, either designed and/or project managed by EES, for the Soccer World Cup 2010 hosted in South Africa.

The five projects are (total construction value shown):

1) Cape Town Stadium – ZAR 4,5 Billion
2) Nelson Mandela Bay Stadium - ZAR 2,1 Billion
3) Cape Town International Airport – ZAR 2,3 Billion
4) King Shaka International Airport – ZAR 7,8 Billion
5) Gautrain Sandton Station – ZAR 1,1 Billion

These projects are outlined in Section 2, followed by a description of the concept of Green ICT in South Africa in Section 3.

To understand sustainability of new buildings, one has to understand the threats to sustainability. These include:

1) Total Cost of Ownership: Most importantly, savings on capital expenditure often translates into high operational expenses of buildings
2) Environmental: With global warming, as well as the high cost of energy, the environmental impact of new buildings is a major concern
3) Revenue generation: Many new buildings are not fully utilized to their potential and opportunities for income generation are missed
4) Workplace effectiveness: Problems with workplace environmental conditions are not always effectively resolved by facilities management
5) Physical and logical security: With many buildings being designed and constructed without ICT value-adds considered, the ICT network and building automation systems may not be as secure as optimally envisaged

These issues are discussed with reference to the above projects in Section 4 and thereafter lessons learned and recommendations are shared in Sections 5 and 6.

2. PROJECT BACKGROUND

This section comprises a brief overview of each of the five World Cup 2010 infrastructures in which EES was involved. Table 1 shows which systems were used for each infrastructure/project.

2.1. Cape Town Stadium (CTS)

EES was the specialist sub-consultant supporting WSP Consulting Engineers at Cape Town Stadium, and was part of the consulting team tasked with designing and implementing the stadium's ICT and BAS infrastructure. The structured cabling system supports the converged active network, which together form the intelligent backbone. An intelligent infrastructure allows the stadium to be managed more efficiently, opens up multi-purpose possibilities and ultimately provides the owner/client with greater value potential. Both Cape Town Stadium and Nelson Mandela Bay Stadium are considered 2010 World Cup showpieces.

2.2. Nelson Mandela Bay Stadium - Port Elizabeth (NMBS)

EES project managed the design and implementation of the intelligent backbone at Nelson Mandela Bay Stadium in Port Elizabeth. All of the stadia's ICT and BAS services operate on the intelligent backbone, demonstrating a true example of convergence. As a multi-purpose stadium it can be quickly transformed from a sporting pitch to rock-concert stage or trade show arena, with minimal adaptation to the lighting, sound, ticketing, safety and management systems from a single operational control centre.

2.3. Cape Town International Airport (CTIA)

Substantial alterations and additions were made to Cape Town International Airport (South Africa's second largest airport) in time for World Cup 2010, breaking new ground regarding cutting-edge efficiency and a far enhanced passenger experience. The new foundation of the entire airport technology infrastructure is now the converged ICT network making the airport an intelligent building in line with world-wide standards.

EES was responsible for project managing the implementation of the Airport's intelligent backbone. This included all the ICT services at the new integrated terminal at the Airport. At the centre of the upgrades and expansion was the new Central Terminal Building (CTB), which opened in November 2009. Work on the CTB involved integrating the domestic and international terminals to streamline operations, boost efficiency, ensure the airport is better utilised and enhance travelers' airport experience.

2.4. King Shaka International Airport – Durban (KSIA)

King Shaka International Airport replaced Durban International Airport as the latter's runway was too short to allow large aircraft such as the Boeing 747 to operate intercontinental routes out of Durban. The new airport now provides for direct long-haul international flights to and from Durban and accommodates the province's rapidly expanding domestic passenger business.

Work at KSIA involved the construction of a fully integrated passenger and freight airport facility. The passenger terminal caters for both domestic and international travel. EES provided technical assistance with the integration, commissioning and operational trial testing of all special systems listed in Table 1.

2.5. Gautrain Sandton Station - Johannesburg

Completion of the Gautrain Sandton Station was achieved in time for the World Cup. Sandton Station is an integral part of South Africa's upgraded, public transport, rapid rail link project, which is still in progress and is currently being extended to further areas.

The link between Sandton and OR Tambo International Airport constitutes Phase One, and it expected that the second phase, linking Johannesburg and Pretoria, will be operational this year. In addition to the three anchor stations on these two links, it is reported that seven other stations will be incorporated in approximately 80 kilometres of rail along the route. EES was integral in coordinating Mechanical, electrical and ICT services for Phase 1.

3. GREEN ICT

Of key importance to all five World Cup 2010 projects discussed in this paper is the need to transform to a greener built environment. The issue of climate change is receiving significant attention the world over. As increasing environmental pressures take hold in SA, the country's major construction companies are striving to pursue green practices and projects.

A Green building is a building constructed in such a way that it mitigates ongoing negative environmental impact. In 2008 the Green Building Council of South Africa (GBCSA) introduced the Green Star SA rating tool (based on internationally recognized green building rating systems), the purpose of which is to set standards and benchmarks for green buildings, and to provide the property development industry with an objective measurement for assessing Green buildings. They provide a concise framework, which when followed and implemented, facilitates 'integrated, entire-building design and construction practices'.

A means to scoring green points, enabling Green Star SA certification, is ICT. Converged network architecture enables stakeholders to obtain building intelligence natively, and then to measure the resulting effectiveness of Greening efforts.

Intelligent network architecture is vital for synchronized automation of the various building systems, such as ventilation, access, safety, power and lighting. And if such a network is in place, the capital expenditure (CAPEX) and operating expenditure (OPEX) of the individual systems could be significantly less than otherwise.

An example of the importance given to Green building and ICT is that of Cape Town Stadium. In the early stages of its development, this stadium was assessed in accordance with the CSIR's Sustainable Building Assessment Tool (SBAT). This tool embraces the triple-bottom-line approach, 'triple' referring to people, profit and planet [1].

Table 1: List of systems in EES's scope for each project. The brief for the Gautrain Project was purely a coordination responsibility (not system specific) and thus is not shown in the table below.

	CTS	NMBS	CTIA	KSIA
Airport Specific				
EVIDS/ BIDS/ FIDS (Electronic Video, Baggage and Flight Information Display Systems)			x	x
CUTE (Common Use TErminal)			x	x
AODB/ ARM (Airport Operational Data Base, Airport Resources Management)			x	x
CUSS (CUstomer Self Service)			x	x
BRS (Baggage Reconciliation System)			x	x
PAS (Personal Identification and Access Control Systems)				x
Aviation Fuel System				x
TETRA (Terrestrial Trunked Radio)			x	x
AGL (Aircraft Ground Lighting)			x	x
BSS (Baggage Sortation System)			x	x
(ADS) Advert Display System			x	
CATV (Community Antenna Television)			x	
AWOS (Automated Weather Observation System)			x	
HBS (Hold Bag Screening)			x	
Loading Bridges			x	
NOC (National Operation Centre)			x	
OCC (Operations Control Centre)			x	
GSM / Cellular			x	
NOC (National Operation Centre)			x	
Stadium Specific/Other				
Wire Centres and IT pre-requisites	x	x	x	x
Voice Over IP / active network	x	x	x	x
CCTV (Closed Circuit TV)	x	x	x	x
BMS (Building Management System)	x	x	x	x
Wifi	x		x	x
Access Control	x	x	x	
Mass Access, Turnstiles and Ticketing	x	x		
Parking	x			x
Support Infrastructure (workstations/ servers/ Uninterrupted Power Supply/ Electronic rooms/ network monitoring)	x	x		
Fire Detection and Evacuation	x		x	
Public Address	x	x	x	
Digital Signage	x	x		
Passive Network (structured cabling)	x	x	x	x
Stadium Big Screens	x	x		
HMI (Human Machine Interface)	x	x		

4. THREATS TO SUSTAINABILITY

4.1. Total Cost of Ownership (TCO)

Fundamental to understanding TCO is CAPEX and OPEX. While energy management systems are seen to be expensive, by using an existing ICT network costs can be reduced.

Cape Town Stadium: A good example is Cape Town Stadium as it has been in operation for one year. The Stadium CAPEX was ZAR 95 million. Usually the OPEX of enterprise environments is 8 to 12% of ICT CAPEX [2], and it is generally accepted that the OPEX of a modern large building should be under 8% of ICT CAPEX. A major contribution to this reduced benchmark is the once off expenditure on structured cabling which if professionally designed and installed should last at least 20 years. Use of a converged network means less maintenance and therefore lower OPEX for first 20 years. In the case of Cape Town Stadium, the ICT OPEX for one year is estimated at ZAR 5.8 million, which is 6.1% of the CAPEX.

Nelson Mandela Bay Stadium: This stadium CAPEX was ZAR 80 million. OPEX for 2011 has been costed by a service provider at ZAR 2.53 million (3.2% of CAPEX). However this has not been awarded yet, as the stadium still has to go out on a formal operational/management company tender. At the moment it is on a cost plus basis on match days.

The above examples demonstrate a reduction in OPEX when compared to the benchmark.

4.2. Environmental

ICT can reduce environmental impact due to the reduction in the materials and energy it utilises.

With reference to the use of materials:

All five projects: The use of converged/structured cabling means less cable is needed than in projects using conventional cabling. GBCSA talks about material use in "Section MAT-10: Dematerialisation" [3] which aims to encourage and recognise designs that produce a net reduction in the total amount of material used. Converged infrastructure enables flexibility and reduces future wastage. Only one campus network is used, which means that there are not numerous separate ICT networks for each of the ICT systems. Less equipment means less power and less cabling.

Cape Town Stadium: ICT enabled the coordination of efficient controls for sub-system implementation e.g. staggered lighting controls. Vendor selection criteria included eWaste policies and capabilities. This helped to drive the roll-out in particular of HP and Cisco take-back programmes through the service provider

With reference to energy savings:

All five projects: All the energy management systems: lighting, generator, power monitoring at MV (Medium Voltage) use the ICT network which means ICT is the enabler, although not a direct contribution. Integrated monitoring and control systems save up to 30% in energy consumption through better management.

Both Stadiums: Energy saving features were incorporated into the initial design. A building management system (BMS) allows for control and monitoring of the active environmental systems, in particular air-conditioning and lighting in different areas, enabling the facilities manager to optimise energy efficiency during operations. The boilers and associated hot water storage are also controlled by the BMS

The parking garage is only fitted with an air supply system in the deep areas away from the perimeter, while the remainder of the parking in this area is naturally ventilated, saving considerable power.

A LAN management system (LMS) is used to turn off switch ports when not in use. This saves approximately 7 Watts per port at an average of 2000 ports = 14kW.

Passive design principles such as 'day-lighting' and natural ventilation are key components that were dealt with during the initial design.

Cape Town Stadium: Through the networked IPTV and the HDMI protocol, all screens can be switched off from a central point using the IPTV software.

There are a multitude of energy savings that are currently in use and not described, the key here is that ICT is an "enabler", it enables the cost effective implementation of a myriad of savings which would otherwise be prohibitive or with lengthy paybacks.

4.3. Revenue Generation

There is tremendous pressure on the stadium owners, sporting bodies or municipalities, for example, to view these assets as revenue generators, rather than loss-making amenities.

Cape Town Stadium: Audio and visual systems installed make the stadium very versatile with regard to the diverse functions and events it can host e.g. concerts like U2 and religious events. It offers 3G and cell coverage, as well as a sophisticated and fully equipped press and media facilities. In general, the intelligent infrastructure allows for most areas of the stadium to be utilized for a variety of functions, office space to conferencing, catering to connectivity and sporting events to live entertainment. Furthermore, multimedia versatility allows for revenue generation via targeted advertising in all areas of use.

Airports and Gautrain: Add on costs to provide services to tenants are marginal, as the head end of the equipment has already been paid for, not to mention that these intelligent infrastructures are more energy efficient than traditional facilities. ICT systems reduce cost of telephony/wifi components etc. Flight Information Display Systems (FIDS) provide advertising opportunities with resulting revenue generation.

Nelson Mandela Bay Stadium: This multi-purpose stadium can be quickly and easily transformed from a sporting pitch to, for example, a concert stage or trade show arena, and this versatility enhances its revenue generating capabilities.

The integrated management system enables stadium management to better manage operations and effectively deploy staff.

General: There is advertising on networked IPTV's throughout all venues, and this makes it very simple and cost effective to enable advertising on top of external broadcaster's multimedia content management.

Venue security is a serious considering factor - the network enables complete security integration, including network controlled turnstiles. Ticketing and security are key contributing factors when determining an events viability.

Networked Point of Sale (POS) terminals offer easy management at lower cost for landlords.

4.4. Workplace effectiveness

To better manage operations and effectively deploy staff, an integrated management system is a benefit to stadium management. They can identify potential problems and deal with them before they become an issue.

Cape Town International Airport and stadiums:
ICT enables intelligent lighting systems, which can automatically simulate natural lighting, proven to be less invasive in an office environment.

The ICT network provides a base for other systems, in turn providing a lower cost for the operation of those systems, making it an enabler for systems installation.

Airports and stadiums: ICT facilitates hot-desking meaning people are able to work from any data point. It also enables HVAC and C02 systems to make it more comfortable for people to work .

4.5. Physical and logical Security

Security needs to be managed via an integrated network to be properly effective e.g. Allan Gray utilizes an integrated system whereby when a person gets removed from the payroll, access-door controllers and all security elements associated will automatically be informed. Fundamental here are physical security databases on a Storage Area Network (SAN) in SQL (database management language).

Both Stadiums: NMBS and CTS offer integrated access control, turnstile and ticketing facilities, meaning that if a person has specific rights to doors, he or she also has access rights to the turnstiles. This can only be achieved if physical and logical security is converged.

An active directory is linked to individual systems, meaning that if an employee vacates their position and is deleted from the active directory he no longer has access to operational applications as his login is linked to the active directory.

5. LESSONS LEARNED

The general lessons learned from the projects are as follows:

1) The ICT engineer should be in control of all major design and implementation documentation - and not simply leave this to the integrator
2) Even if you are taking over the project from another company or organisation, start again – especially time sensitive projects, don't waste time trying to figure out what they did
3) The builder must co-ordinate, control and quality check all his sub-contractors' installations
4) Co-ordination issues can severely affect response and action to programme timeframes
5) Co-ordination communication between Civils & Building and Mechanical & Electrical is critical
6) It is important to gain client appreciation of the time it takes to commission the building services portion of a project
7) It is vital to have the client 100% behind you - suspicion/resistance from clients towards the team is problematic
8) Appoint an operator/ICT manager early on. Late appointment of an operator means that the integrators and professional team are required to manage and supply additional requirements, resulting in delays and further costs. In this situation the consultant has had to make technical decisions based on 'assumed' operational requirements during design, construction and commissioning phases
9) Choose an integrator likely to carry you into the future. In making this choice the visibility of the integrator resource and expertise base is important; if this information it is unknown it can be very risky to the project
10) ICT is not traditionally considered part of the construction phase. This effectively means that upfront planning is not performed adequately and the stakeholders or users of the network are not always considered at the right times. It also means that we do not have an effective procurement model to enable ICT to be procured during the construction phase

6. RECOMMENDATIONS

The implementation of an intelligent backbone must have a team approach and all major players should integrate the systems. Experts should co-operate and make use of their own particular strengths to deliver a complete solution.

Using a silo approach (each system considered separately) contributes to a disconnect between services, which makes it complicated for owners and visitors to the structures to take full advantage of the benefits of intelligent building systems, such as energy monitoring, security systems, and integrated communications. A Master Systems Integrator that implements the total ICT solution is often favourable for fully converged systems.

Achieving on intelligent backbone entails the installation of both ICT and BAS. Planning must happen upfront. ICT infrastructure should be included in the early planning phase of stadia design and development to ensure that the ICT and BAS are designed to support current and future information needs. There must be sufficient flexibility in the design of the intelligent backbone to allow for future inclusion of feature-rich services that were perhaps not initially envisaged or available at the time of planning. It must support not only today's technology, but tomorrow's as well.

Convention shows that developers listen to architects and consultants. It's our experience that this is where the principal problem exists. Very few such service providers have the skills and knowhow to incorporate an integrated ICT/BAS concept into the building design .

When investing in infrastructure, it is vital that sustainability is entrenched in the construction design from day one. Technology companies can help the construction industry to achieve this. Today a new thinking in the approach to projects is needed. Stakeholders need to listen to each other and form a working partnership in the early days of a project. This will benefit future generations and positively impact into the global built environment, going forward.

7. CONCLUSION

Based on their project experience, it is possible for an intelligent building design team to improve building sustainability by setting energy performance goals and managing the building system's inter-dependencies culminating in the production of a more efficient, cost-effective and smart building. A master systems integrator is recommended to implement the design.

Sustainability should be a key focus area for developers and other stakeholders. ICT can be a significant enabler in achieving sustainability.

Its apparent though that a major obstacle facing the ICT industry is the human tendency to resist change. The individual system silo approach is still favoured because it is familiar and perceived as easier to implement, compared with the 'unfamiliar' advantages and benefits of converged technologies.

In closing, the current and future built environment professionals need to be educated and up-skilled by the ICT industry, in order to understand the benefits. This will ensure that there is no resistance to a dialogue between professionals at the conception phase of a project thereby ensuring that the required infrastructure is installed and the above mentioned threats to building sustainability are overcome.

8. REFERENCES

[1] GREEN by DESIGN WSP, PJ Carew Consulting, CSIR, (2010), *2010 FIFA World Cup Green Point Stadium environmental performance enhanced*, Brooke Pattrick Publications

[2] http://www.gartner.com/technology/research/enterprise-it-leaders.jsp [accessed 2 February 2011]

[3] Green Building Council of South Africa, (2008) *Technical Manual Green Star SA Office Design and Office As built,* Version 1.

Electrochemical Treatment for PAHs Degradation in Virgin and Recycled Non-Ionic Surfactant Solutions

Zainab Z. ISMAIL
Dept. of Environmental Engineering, University of Baghdad
Al-Jaderia, Baghdad, Iraq

and

Taghreed A. HUSSEIN
Dept. of Environmental Engineering, University of Baghdad
Al-Jaderia, Baghdad, Iraq

ABSTRACT

The removal of carcinogenic polycyclic aromatic hydrocarbons (PAHs) from soils presents a challenge to scientists and engineers. The high hydrophobic nature of PAHs enables their strong sorption onto soil particles. Thus, the application of surfactant washing systems may be a favorable option to release these pollutants from contaminated soils. One of the main limitations for the surfactants applications in soil remediation is the lack of knowledge about the environmental fate of the PAH-loaded surfactants. The investigation presented here focused on the electrochemical destruction of anthracene and pyrene extracted from contaminated soils by two non-ionic surfactants, Triton X-100 and Tween 80. The electrochemical treatment of these solutions was carried out in a glass cubic electrolytic cell of 0.4L working volume and graphite electrodes material. Nearly complete degradation of model PAHs was reached for all the experiments

Keywords: Soil pollution, Soil washing, Electrochemical treatment, PAH-destruction and Surfactant recycling.

1. INTRODUCTION

A potential technology for rapid removal of PAHs adsorbed to soils is washing the contaminated soil with surfactant solutions for the ability of these chemicals to partition hydrophobic organic compounds into their micelle core. Surfactant-enhanced soil remediation is an excellent alternative to pump-and-treat for remediation of PAHs-contaminated soil [1]. One of the main limitations for the surfactants applications in soil remediation is the lack of knowledge about the environmental fate of the PAH-loaded surfactants. Once PAHs enter into wastewater, it is difficult to remediate by conventional water treatment methods, as they are essentially recalcitrant, persistent and non-reactive in water [2]. Accordingly, many research efforts have been expanded to find suitable methods for remediation of soil and water environments contaminated with PAHs [3]. Solvent extraction, activated carbon adsorption, and ion exchange can strip the contaminant out of surfactant micelles leaving the micelles intact in solution [4]. Among these methods, an environmentally friendly approach for PAH degradation in aqueous solution could be based on the electrochemical treatment [5].
The objective of the present study is to evaluate the validation of reusing the electrochemically treated surfactants for excessive cycles of soil washing process.

2. MATERIALS AND METHODS

Materials

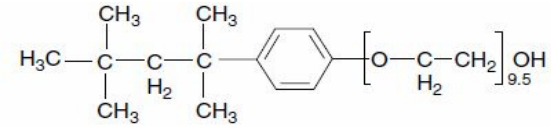
Two types of non-ionic surfactants, Triton X-100 ($C_{34}H_{62}O_{11}$) supplied by Thomas Baker (India) and Tween 80 ($C_{24}H_{44}O_6(OCH_2CH_2)_{20}$) obtained from Atlas Chemical Company (England), were used to carry out this study. Each surfactant solution was previously used for alternatively extracting anthracene and pyrene from PAHs-contaminated soil samples.

The characteristics of the surfactants are given in Table1

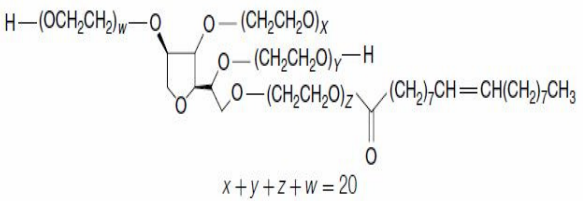
Table 1 Properties of the surfactants as provided by the supplier.

Surfactant	Mwt (g/mol)	HLB	Purity
Triton X-100	646.87	13.5	99%
Tween 80	1309.0	15.0	97%

The chemical structure of the surfactants and the model PAHs used in this study are given in Figs.1 and 2.



Triton X-100



Tween 80

Fig.1 Chemical structure of the tested non-ionic surfactants.

Anthracene **Pyrene**

Fig.2 Chemical structure of the model PAHs.

Electrochemical cell

For this study, an electrochemical glass cell of 0.4 L working volume was used. The cell had a cubic body with a working graphite electrode immersed area of 52 cm^2 and an electrode gap of 8 cm. Sodium sulfate (Na_2SO_4) was selected as an inert supporting electrolyte of 0.1 M concentration.

All experiments were conducted at constant voltage drop (5V) applied through a couple of graphite electrodes with a DC power supply (Model L30E). Magnetic stirring was used to prevent concentration and pH gradients. The experimental set-up is given in Fig.3

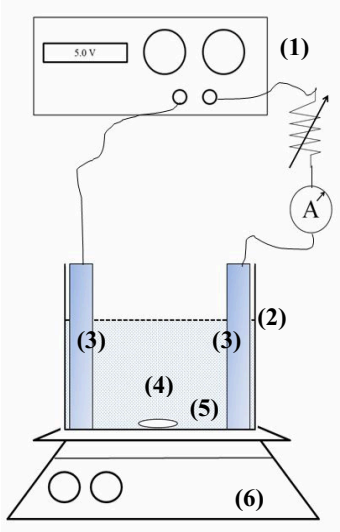


Fig.3 Experimental set-up: (1) electric power supply, (2) electrolytic cell, (3) graphite electrodes, (4) PAHs-loaded surfactant solution, (5) magnets, (6) magnetic stirrer.

Experimental procedure

The PAHs-loaded surfactant solutions were treated by electrochemical technique to destroy the previously extracted PAHs from contaminated soil samples and to allow recycling of the PAH-free surfactant solutions for excessive soil washing cycles. Four different surfactant solutions (Table 2) were alternatively treated in the electrolytic cell.

Table 2 PAHs concentrations in the non-ionic surfactant solutions.

Sol. No.	Surfactant	PAH	PAH conc. (mg/L)
1	Triton X-100	Pyrene	71.63
2	Triton X-100	Anthracene	49.39
3	Tween 80	Pyrene	58.27
4	Tween 80	Anthracene	41.17

The difference in the initial concentration of the model PAHs in the surfactant solutions is due to the different desorption capacities of pyrene and anthracene from the spiked-soil as it was in the order of pyrene > anthracene.

Samples of the electrochemically reacted mixture were taken from the electrolytic cell

every 2 h to be analyzed by HPLC for determination of PAHs concentrations gradient. The electrolyte pH and conductivity were monitored along the treatment period.

In order to estimate the validity of the electrochemically treated surfactant solutions, these solutions were reused for extracting alternatively pyrene and anthracene from PAHs-spiked soil samples. The initial concentrations of pyrene and anthracene in soil samples were 500 mg PAH/kg soil. Then after, the PAHs-loaded recycled solutions were retreated electrochemically to destroy the extracted PAHs.

3. RESULTS AND DISCUSSION

In order to obtain PAHs-free washing solutions, the PAH-loaded surfactant solutions were treated for complete degradation of PAHs. In the present study, electrochemical oxidation was proposed as a convenient treatment for PAHs destruction in which, the electrical current induces redox reactions upon the surface of the electrodes and the oxidation reaction results in the destruction of the refractory organic pollutants. The time-courses of anthracene and pyrene destruction in virgin solutions of Triton X-100 and Tween 80, respectively are shown in Fig. 4.

According to Tran et al. [6], the anodic oxidation of pollutants occurs heterogeneously. Initially, the organic pollutants will transport toward the anode electrode surface to be oxidized, and then two different regions can be recognized as illustrated in Fig. 4. In the first region, the degradation of PAHs increases almost linearly within the initial duration. Beyond 10 h, the rate of PAH degradation remains nearly stable.

The decomposition of the pollutants occurs due to electrophilic attack of oxidizing species, e.g. hydroxyl radicals, ozone which are generated electrochemically. These species generated from destruction of water molecule at anode surface and react favorably with electron-rich compounds. They efficiently react with the carbon double bonds C=C and attack the aromatic nucleus, which are the major component of refractory PAH.

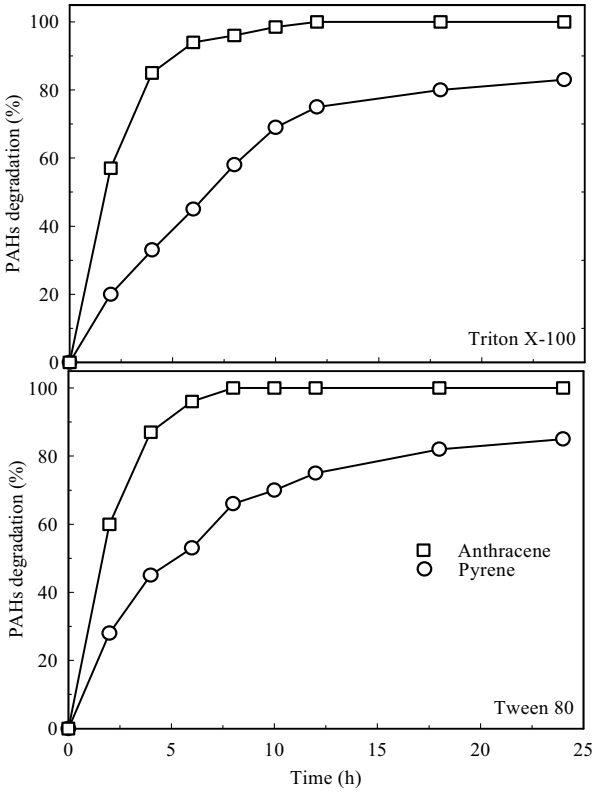


Fig.4 Profiles of PAHs electrochemical degradation with time in the virgin surfactant solutions.

Similar profiles were obtained for the degradation of anthracene and pyrene in Triton X-100 and Tween 80. The initial slope of anthracene degradation was higher, reaching an average value of about 95% in 6 h, increasing up to complete degradation after 10 h. For pyrene, it is notable that the initial and final degradation levels were relatively lower than anthracene. This could be attributed primarily to the level of aromatic fractions of the model PAHs molecules (Fig.2) as higher fraction of aromatic fractions may need longer time to be destructed. Also, the higher level of pyrene concentration (Table 2) may represent another potential reason for the lower degradation rate of pyrene compared to anthracene.

However, for anthracene and pyrene, their degradation rate in Triton X-100 was slightly lower than in Tween 80 and this result may be explained by the different hydrophobicity of the surfactant solutions which in turn could affects the interaction between the miceller core and certain PAH. Thus, in case of anthracene, after 24 h of electrochemical treatment, complete degradation up to 100% has been obtained. For

pyrene, few more hours up to 6 h may be needed to obtain a complete degradation. Such a fact points out the potential of this technique for PAH degradation in surfactant solutions as well as being compatible with the environment because the main reagent which is the electron is a clean one [7].
Additional observations can be estimated relative to the electrochemical degradation of PAHs as follows:

● The PAHs degradations didn't affect the pH of the electrochemically treated surfactant solutions since the pH remained in the range of 7.4-7.9 in the electrolytic cell during the progress of the PAHs degradation process. Preventing concentration and pH gradient was due to the continuous mixing of the electrochemically treated PAH-surfactant solution using magnetic stirring.

● Prior to adding the inert electrolyte Na_2SO_4, the electrical conductivity values of the surfactant solutions in the electrochemical cell were at a range of 31.6-38.8 ms/cm for the two types of PAHs in each surfactant solution. Upon the addition of Na_2SO_4, consequently electrical conductivity increased up to about 65 ms/cm. This raise in the electrical conductivity is well expected [8], which in turns increased the reaction rate significantly, and almost complete oxidation of the PAHs was reached in shorter treatment time with adequate electric power consumption.
The percentages of PAHs destructions in the electrochemically treated surfactant solutions at 10 and 24 h are given in Table 3.

Table 3 Electrochemical destructions of PAHs in virgin surfactant solutions at 10 and 24 h.

PAH	PAH destruction (%)*	
	10 h	24 h
	Triton X-100	
Anthracene	98.5	100
Pyrene	69.0	83.0
	Tween 80	
Anthracene	100	100
Pyrene	74.0	89.3

*Average of 2 values.

Due to the relatively high cost of surfactants, it is necessary to examine the possibility of recycling the washing solution and to verify the efficiency of this solution in a successive soil washing process. The percentage removal of PAHs from soil samples using recycled surfactant solutions compared to virgin solutions are given in Table 4. These promising results indicate the validity of reusing the electrochemically treated surfactant solutions for subsequent soil washing cycle.

Table 4 Comparison between the efficiency of PAHs desorption from soil by virgin and recycled surfactant solutions.

PAH	PAH desorption (%)	
	Virgin solution	Recycled solution
	Triton X-100	
Pyrene	86.3 ± 2.6*	83.8 ± 3.5*
Anthracene	59.5 ± 4.1	57.5 ± 6.3
	Tween 80	
Pyrene	70.2 ± 4.6	68.4 ± 3.8
Anthracene	49.6 ± 2.8	47.5 ± 3.4

* ± represent one standard deviations; n = 3.

Upon completion of soil washing cycle with the recycled surfactant solutions, they re-subjected to an electrochemical treatment for the destruction of newly extracted PAHs. The electrochemical degradation profiles of anthracene and pyrene in recycled surfactant solutions are presented in Fig.5. The profiles and the rate of PAH degradation are nearly similar to those obtained as for PAHs destruction in virgin surfactant solutions (Fig.4). The percentages of anthracene and pyrene destruction in the recycled surfactant solutions are presented in Table 4.

Table 4 Percentages of PAHs electrochemical destruction in recycled surfactant solutions at 10 and 24 h.

PAH	PAH destruction (%)*	
	10 h	24 h
	Triton X-100	
Anthracene	98	100
Pyrene	64	83
	Tween 80	
Anthracene	100	100
Pyrene	70	85

*Average of 2 values.

As shown in Tables 3 and 4, the degrees of PAHs degradation in recycled solutions are approximately at the same level as in virgin surfactant solutions.

These findings suggest a potential approach for PAHs degradation due to the high efficiency of electrochemical treatment as well as the flexibility of recycling the surfactant solutions for excessive treatment cycles.

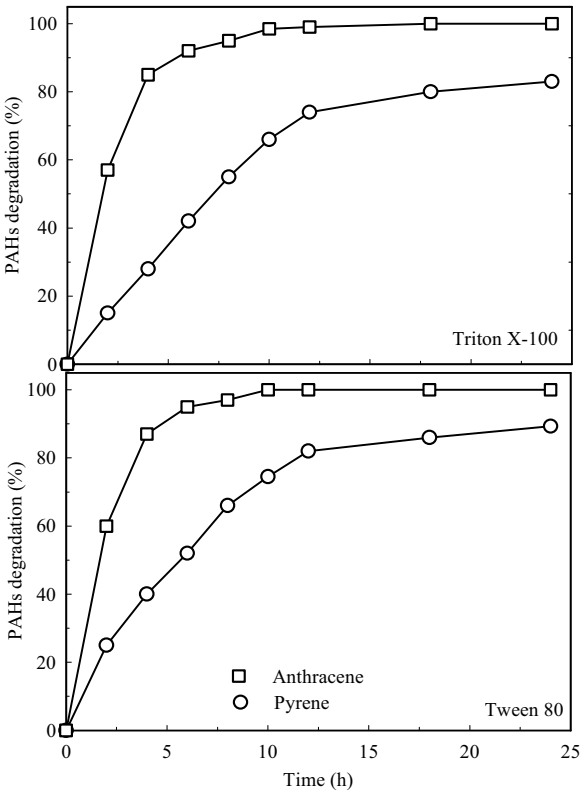


Fig.5 Profiles of PAHs electrochemical degradation with time in recycled surfactant solutions.

4. CONCLUSION

This study has demonstrated the effectiveness of the electrochemical treatment employed for PAHs degradation in virgin and recycled micellar forming suspensions. The remediation of the PAHs-loaded surfactant solutions has been developed and evaluated at a laboratory scale in a cubic glass electrolytic cell comprising of graphite electrodes. The liquid collected from the electrokinetic remediation process, which was polluted with anthracene or pyrene was successfully oxidized by electrochemical treatment ready to be recycled for excessive soil washing cycles. The quality of the recycled surfactant solution and their efficiency to clean-up the PAHs-contaminated soil were approximately similar to the virgin solutions quality indicating the potential of the electrochemical treatment application.

It can be concluded that the electrochemical treatment described here was highly suitable for PAHs degradation.

5. REFERENCES

[1] H. Cheng, D.A. Sabatini, "Simultaneous uptake of anionic surfactants and micellar solubilized contaminants using anion-exchange resins", **Water Research**, Vol. 36, 2002, pp. 2062-2076.

[2] Y. Gao, W. Ling, M.H. Wong, "Plant-accelerated dissipation of phenanthrene and pyrene from water in the presence of a nonionic-surfactant", **Chemosphere**, Vol. 63, 2006, pp. 1560-1567.

[3] M.T. Alcántara, M. Gómez, M. Pazos, M.A. Sanromán, "Combined treatment of PAHs contaminated soils using the sequence extraction with surfactant-electrochemical degradation" **Chemosphere,** Vol. 70, 2008, pp.1438-1444.

[4] C.K. Ahn, Y.M. Kim, S.H. Woo, J.M. Park, "Soil washing using various nonionic surfactants and their recovery by selective adsorption with activated carbon", **Journal of Hazardous Materials**, Vol. 154, 2008, pp.153-160.

[5] M.T. Alcántara, J. Gómez, M. Pazos, M.A. Sanromán, "PAHs soil decontamination in two steps: Desorption and electrochemical treatment", **Journal of Hazardous Materials**, Vol. 166, 2009, pp.462-468.

[6] L.H. Tran, P. Drogui, J.F. Blais, G. Mercier, "Electrochemical degradation of polycyclic aromatic hydrocarbon in creosote solution using ruthenium oxide on titanium expended mesh anode", **Journal of Hazardous Materials**, Vol. 164, 2009a, pp. 1118–1129.

[7] K. Jüttner, U. Galla, H. Schmieder, "Electrochemical approaches to environmental problems in the process industry",

Electrochimica Acta, Vol. 45, 2000, pp. 2575–2594.

[8] C. Cameselle, M. Pazos, A. Sanroman, "Selection of an electrolyte to enhance the electrochemical decolourisation of Indigo. Optimization and scale-up", **Chemosphere,** Vol. 60, 2005, pp.1080–1086.

Decision Mapping and Optimal Inspection Models for Plant Maintenance: Some Case Studies

Prof. Dr. K. P. Ramachandran
Associate Dean (PG&R)
Caledonian College of Engineering,
P.O. Box 2322, CPO 111 Seeb, Sultanate of Oman

and

Ahmed Fadhil Said Al Hinai
Oman Refineries and Petrochemicals Company LLC,
P.O. Box 3568, CPO 112 Ruwi, Muscat, Sultanate of Oman

ABSTRACT

The need for an effective maintenance strategy and management system has been recognized as a very important and critical for reducing operational and maintenance costs, especially when the equipment increases in size and complexity. Selection of an appropriate and adequate maintenance and its frequency is considered to be crucial for failure free operation of plant equipment maintenance in process industries, particularly in desalination plants. In this paper a Decision Making Grid (DMG) technique has been applied to select the most suitable type of maintenance. Further, this paper outlines mathematical models that are applied to determine the optimum inspection frequency based on stochastic failures under preventive maintenance. Mean time between failures, mean time to repair and mean inspection time are fitted to compute optimum inspection frequency under preventive maintenance. Case studies from a desalination plant are presented based on the actual break down details and maintenance information.
Keywords: Mean Time between Failures, Mean Time to Inspections, Mean Time to Repair, Condition Based Maintenance, Preventive Maintenance, Decision Making Grid .

1. INTRODUCTION

Modern production equipment has been designed with greater production capacity and higher reliability to meet the rigorous operating demands of industries, especially process industries. These improvements have also made the equipment more complex and challenging to maintain. Over the past few years, there is enormous emphasis to improve the maintenance performance. However if maintenance strategy is to be effective, it must be supported with a decision on adequate resources and maintenance information. Computerized maintenance system and decision models provide the maintenance planners to take decisions on how a plant should be maintained and how often and how much maintenance is to be done in an effective manner with minimum cost [1].The application of various maintenance strategies during the life period could be established based on the failure pattern of any system. In Operate To Failure (OTF) or Breakdown Maintenance, machine is repaired once it fails and the downtime expected here is short. Fixed Time Maintenance (FTM) or Preventive Maintenance (PM) is a time-based strategy where maintenance actions are performed on a pre-determined, periodic basis. PM involves the repair, replacement, and maintenance of equipment in order to avoid unexpected failure [2]. The objective of any PM program is the minimization of the total cost of inspection and repair, and equipment downtime (measured in terms of lost production capacity or reduced product quality). The traditional approach in PM is based on the use of statistical and reliability analysis of equipment failure. Under statistical-reliability (S-R) based PM, the minimum total cost objective is pursued by establishing a "optimal" PM interval at which the equipment or system can be replaced or overhauled [3].
Condition Based Maintenance (CBM) approach involves the use of sensor-based monitoring of equipment condition in order to predict machine failure. Under condition-based maintenance, intervals between PM works are no longer fixed, but are performed only "when needed" [4]. Design Out Maintenance (DOM) and Skill Level upgrade (SLU) are also well considered maintenance strategies in industries. In DOM, the equipment is designed for minimum maintenance, but takes a long time to bring it back upon failure, whereas the machine can be repaired very fast upon failure in SLU.

2. DECISION MAKING GRID

Grid Analysis or Decision Making Grid (DMG) is a useful technique for making a decision on maintenance and considered to be powerful where a number of alternatives are available to choose from, and many different factors to take into account [5]. The DMG map has different levels/ranges as rows on a table, and the factors/criterion of maintenance as columns. The score of each option/factor combination is weighted and added these scores are accumulated to give an overall score for the option. DMG is an effective method that uses a multiple criteria for the evaluation of machine performance such as downtime and frequency of failures. DMG model is very effective in getting rid of critical problems of machines. Further, DMG deals with the selection of the appropriate maintenance strategy that is suitable for prioritizing failure analysis. When there are several options, DMG tool helps the maintenance engineer to choose the correct decision amongst other different factors [6].

3. MATHEMATICAL MODELS

3.1 Inspection models

The availability A of an equipment or plant is given by a simple equation

$$A = \frac{R}{R + M} \quad (1)$$

Where, R is the Reliability and M the Maintainability. MTBF(Mean Time Between Failures) is a measure of the reliability and is fixed for given equipment, whereas MTTR (Mean Time To Repairs) is a measure of maintainability[7]. The availability of machine can be increased by reducing the MTTR which could be possible only by an effective and meaningful maintenance strategy. The maintenance decision is primarily related to inspection of machine for finding the working condition. The inspection of machine can be optimized based on past information on the machine and their interpretations. Decision on optimal inspection could be complicated as it entails consideration of cost, down time, production demand, preventive maintenance shut down and survival time etc. Further, for a constant failure rate (λ) of a machine, the following simple relation for the examination interval can be deduced [4]:

$$e^{\lambda t} - \lambda t = 1 + \lambda \frac{C_i}{C_R} \quad (2)$$

where c_i is the examination cost and c_R is the unitary downtime cost or cost of repair.

Our goal is to determine the inspection policy which minimizes the total downtime per unit time incurred due to a breakdown as well as to maximize the profit per unit time. The total down time D(n) is considered as sum of down time incurred due to repairs per unit time and inspection per unit time[8]. This will give rise to equation.3

$$D(n) = \frac{\lambda(n)}{\mu} + \frac{n}{i} \quad (3)$$

Where μ is the Mean Service rate= 1/MTTR and i is the Mean inspection rate =1/Mean Time to Inspection (MTTI). Differentiating above equation,

$$\frac{d}{dn}[D(n)] = \frac{\lambda'(n)}{\mu} + \frac{1}{i} \text{ and hence } \lambda'(n) = \frac{-\mu}{i} \quad (4)$$

Substituting for $\lambda(n) = \frac{K}{n}$ and $\lambda'(n) = \frac{K}{n^2}$; where the K can be interrupted as the arrival rate of break-downs per unit time when one inspection is made per unit time. We obtain an expression for inspection frequency to minimize the down time as

$$n(D) = \sqrt{\frac{i\,K}{\mu}} \quad (5)$$

We can extendequation.5 for maximization of profit as

$$n(P) = \sqrt{\frac{i\,K}{\mu}\left[\frac{C_u + C_R}{C_u + C_i}\right]} \quad (6)$$

Where C_u is the profit of the value of the output in an uninterrupted unit of time if there is no downtime losses, C_i is the average cost of inspection.

3.2 Preventive Maintenance replacement models

There are two basic preventive maintenance policies for the systems that are subject to stochastic failure. They are age-based replacement policy and constant-interval replacement policy [9].

According to age-based replacement policy, preventive replacement is performed after t_p hours of continuous operation without failure. If the system fails prior to t_p hours, maintenance (replacement) is performed at the time of failure and preventive maintenance is rescheduled after the same t_p operational hours.
In a constant-interval replacement policy, preventive replacement is performed after it has been operating for the total t_p hours, regardless of the number of intervening failures. If a failure occurs prior to t_p hours, only a minimal repair is performed. Minimal repair does not change the failure rate of the system and preventive maintenance (replacement) renews the system and it becomes as good as new.

An age-based replacement policy is generally suited for simple equipment or single-units in which repair at the time of failure (replacement) corresponds to a general overhaul. A constant-interval replacement policy is suited for complex systems like engines and turbines which has a larger number of units within itself. In this paper, the constant-interval replacement policy is reviewed to determine the optimal time for preventive maintenance, t_p, as age-based replacement policy is relevant to simple equipment only [10].

Total expected cost of preventive and breakdown maintenance per unit time,

$$\mathrm{UEC}(t_p) = \frac{\text{Total expected cost due to PM \& minimal repairs}}{\text{Expected length of interval}}$$

$$= \frac{C_p + C_R H(t_p)}{t_p} \quad (7)$$

Where, $H(t_p) = \int_0^{t_p} \lambda(t)\,dt$ $\lambda(t) = \frac{f(t)}{R(t)}$ and C_R is the unitary downtime cost or cost of repair and C_p is the cost of PM replacement.

And f (t) is Time-to-failure probability density function, R(t) is the reliability and $t_p$is Time for preventive replacement. The optimal $t_p$is the value of $t_p$that minimize the function $\mathrm{UEC}(t_p)$.

4. CASE STUDIES:

4.1 Decision Making Grid

In this paper, the selection of appropriate maintenance strategy for a desalination plant is devised using DMG. Approach adopted here is categorizing the desalination plant failures by considering their downtimes and frequency of failures. Various factors are taken into account while developing DMG model such as machine criticality, downtime, frequency of failures, machine age and other operational limitations.

The following steps are used to implement the DMG for the desalination plant.

Step.1: Criteria Analysis**:** In this step, an attempt is made to assess how badly the component is performed for a period of two years' time. The worst performers are sorted and placed into high, medium, and low sub groups according to number of failures and down time as shown in Table.1. It is a priority table

for the failures based on down time and number of failure of an item.

Step.2: Decision Mapping: The aim of this step is to monitor the worst component of the desalination unit on a grid as shown in Figure.1. This is done by assigning the type of maintenance to be followed depends upon the down time. Here the grid acts as a map on which the performance of worst components is located according to multiple criterions.

Table.1
Data Analysis for the DMG

	Name	Down Time (hours)	Name	Frequency	
High	Tube Leaks	48	Tubes Leaks	15	**High**
	Bundle Fouling and Scaling	48	Bundle Fouling and Scaling	25	
	Screen Jetting	48	Screen Jetting	6	
	Filters(Demister)	48	Gaskets	5	
	Chemical Air Pumps	24	Feed Pump	5	
	Feed pumps failure	24	Chemical Air Pumps	5	
Medium	Gaskets	12	Pin Hole Leaking	5	**Medium**
	Pin Hole Leaking	12	Vales Failure	4	
	Valve Failure	12	Loss of Vacuum	2	
	Temperature Gauges	12	Temperature Gauges	2	
	Chemical Drum Leaking	12	Bolts broken	2	
Low	Loss of Vacuum	6	Chemical drum Leaking	1	**Low**
	Instrument line failure	6	Instrument line failure	1	
	Sample point clogging	6	Sample point clogging	1	
	Bolts broken	6	Filters (Demister)	1	
	Broken knobs of pumps	6		1	
	Loosen of different types of flanges	3.6	Loosen of different types of flanges	1	

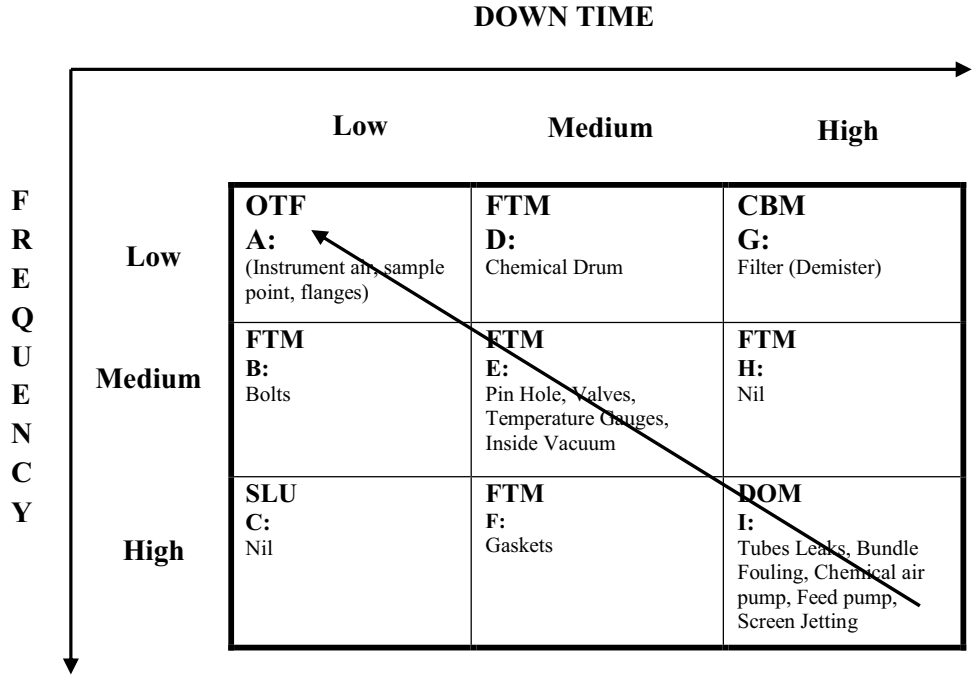


Figure 1: Decision Mapping

Step.3: Multi leveled decision support: Once the worst performing machines are identified and the appropriate action is suggested, it is now moved from strategic systems level to the operational component level [11]. Figure.2 represents criteria of evaluation (objectives), types of machine, failure categories and failure details etc. The criteria evaluation is based on downtime (in minutes or months), frequency of failure, spare parts requirement and bottlenecks in work execution. Further, these criteria are mapped under various failure categories and then fault details are assimilated. Interrelationship between the criteria and failure modes are indicated by arrows. Table.2 depicts the DMG obtained from a multilevel decision support based on Figure 2.

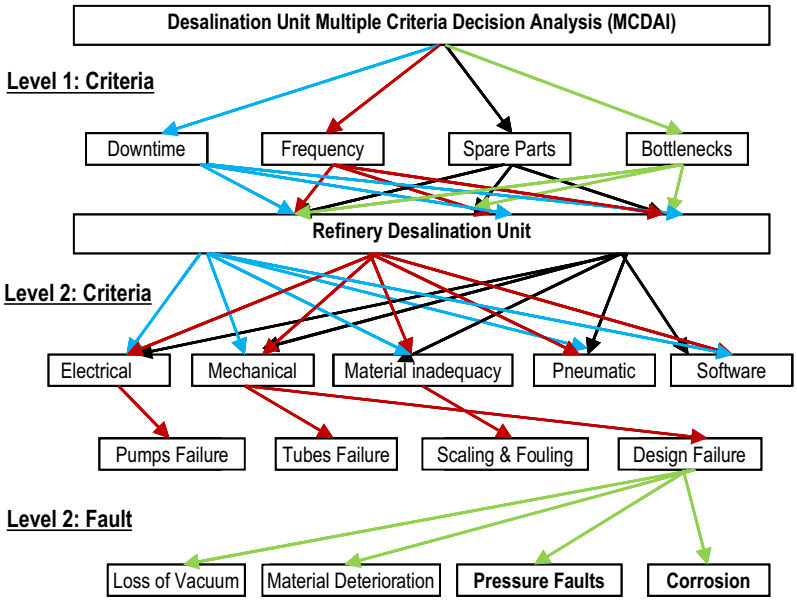


Figure 2: Multi leveled decision support

Table.2
Decision Making Grid

Sl No	Downtime	Frequency	Failure Name	Decision taken
1	Low	Low	Instrument air, sample point, flanges, Pumps broken knobs	OTF
2	Low	Medium	Bolts	FTM
3	Low	High	Nil	SLU
4	Medium	Low	Chemical Drum	FTM
5	Medium	Medium	Pin Hole, Valves, Temperature Gauges, Desal Inside Vacuum	FTM
6	Medium	High	Gaskets	FTM
7	High	Low	Filter (Demister)	CBM
8	High	Medium	Nil	FTM
9	High	High	Tubes Leaks, Bundle Fouling, Chemical Air Pump, Feed Pump, Screen Jetting	DOM

The DMG model developed and proposed for implementation could support the decision making on strategies of maintenance for equipment in the desalination plant.

4.2 Inspection models

Maintenance records were reviewed and the breakdown and maintenance information are obtained for determining optimum inspection frequency. The optimum inspection frequency based on cost and down time computed using the model described by equation.5 and 6 is exhibited in Table.3. Here, the inspection cost per month and cost of products per month (production loss) are taken as RO (Rial Omani) 2.750and 20910 respectively.

Table.3
Optimum Inspection Frequency

SL.NO	Inspection items	K	MTTR (Minutes)	MTTI (Minutes)	Cost of Repair in RO	Optimal Inspection	
						n(P)	n(D)
1	Tube Leaks	0.625	192	30	3251	2	2
2	Bundle Fouling and Scaling	1	120	60	114	2	2
3	Screen Jetting	0.25	480	15	179	3	3
4	Filters	0.042	288	30	9.6	1	1
5	Chemical Air pumps	0.21	288	30	52	3	3
6	Feed Pumps Failure	0.21	288	60	62	3	3
7	Gaskets	0.21	144	30	27	1	1
8	Pin Hole Leaking	0.125	240	30	19	1	1
9	Valves failures	0.21	144	60	51	1	1
10	Chemical drum Leaking	0.042	720	15	29	1	1
11	Bolts broken	0.083	180	30	10	1	1
12	Instrument failure	0.042	360	30	22	1	1
13	Sample point clogging	0.042	360	40	7.5	1	1
14	Loss of vacuum while starting	0.083	180	60	15	1	1
15	Pumps Control	0.042	360	15	12	1	1
16	Temperature Gauges	0.125	120	15	27	1	1
17	Flanges	0.042	216	20	6	1	1

4.3 Replacement models

The breakdown details of the three similar sea water pumps for past three years (36 month) were studied. It is assumed that during three years this equipment follows a uniform time to failure probability density function. Considering pump.1, the cost of preventive replacement is RO 337 and the cost of failure replacement is RO 1500, the optimal time for preventive replacement can be estimated as follows,

$$f(t) = \frac{1}{36} \quad 0 \le t \le 36 \tag{8}$$

$$R(t) = 1 - \frac{1}{36}t \tag{9}$$

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{36}}{1 - \frac{1}{36}t} = \frac{1}{36 - t} \tag{10}$$

$$H(t_p) = \int_0^{t_p} \lambda(t)\,dt = \int_0^{t_p} \frac{1}{36 - t}\,dt = \ln\frac{36}{36 - t_p} \tag{11}$$

$$UEC_p(t) = \frac{C_p + C_R H(t_p)}{t_p} = \frac{337 + 1500 \ln\frac{36}{36 - t_p}}{t_p}$$
$$= \frac{337}{t_p} + \frac{1500}{t_p} \ln\frac{36}{36 - t_p} \tag{12}$$

The expected cost of PM is computed for different values of time as depicted in Table.4.The value of t_p that gives the lowest expected cost (UEC) is shown in Figure.3.

Table.4
Expected cost of PM

Cost	Cost of failure replacement (RO)	Cost of Preventive Maintenance (RO)	Time t_p in months															
			10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Pump 1	1500	337	82	80.3	78.7	77.6	76.8	76.3	**76.1**	76.2	76.5	77	77.7	78.6	79.7	81.1	82.7	84.6
Pump 2	1700	460	102	98.7	96.2	94.4	93.1	92.2	91.6	**91.3**	91.7	91.7	92.2	93.1	94.2	95.5	97.2	99.3
Pump 3	1100	515	87.3	83.3	80	77.5	75.5	73.9	72.6	71.6	71	70.5	**70.3**	70.4	70.6	71.1	71.8	72.8

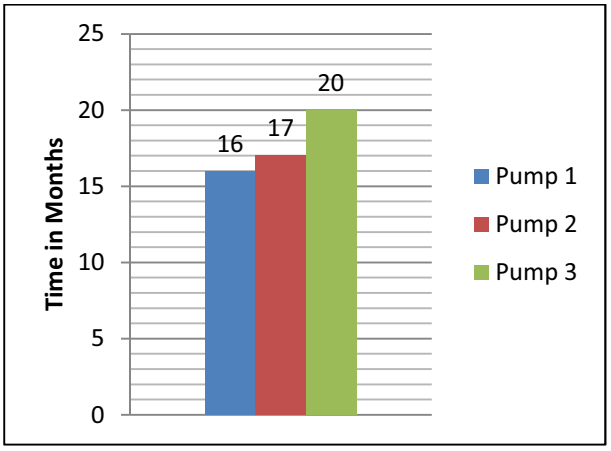


Figure 3: Optimal replacement time

5. CONCLUSION

Sea water desalination is extremely vital for the daily life activities and for the development of GCC states like Sultanate of Oman. The desalination plant has been suffering from frequent shutdowns and repairs over the past years. It is apparent that there is a lack of a proper maintenance strategy which could reduce down time and unwarranted failures. Based on the analysis that was carried out, a DMG model has been developed for the selection of appropriate maintenance strategies in order to reduce the frequency of failures and the downtime of the desalination plant. The DMG model developed is very effective in getting rid of critical problems of machines and to select appropriate maintenance strategy for the Desalination plant considering various aspects such as down time, frequency of failure, spare parts, failure modes etc. A fair estimate of inspection frequency has been derived considering MTTR, MTTI, breakdown details and various cost elements. Implementing the inspection models obtained can reduce the frequency of failures and will prolong useful operating time of plant equipment. Further, the mathematical model has been applied to determine optimum replacement for seawater pumps under constant interval replacement policy. The result shows that the optimum PM replacement is influenced by failure rate, cost of replacement and cost of PM instead of a fixed schedule.

REFERENCES

[1] Lawrence Mann Jr, Anuj Saxena and Gerald M.Knapp, "Statistical-based or condition-based preventive maintenance?", **Journal of Quality in Maintenance Engineering.** 1(1), 1995, pp. 46-59.

[2] Anthony M Smith and Glenn R Hinchcliffe, "Maintenance as a profit center takes a change in thinking", *Plant Engineering;* Sep 2005, **ProQuest Science Journals**, 2005, pp. 28-32.

[3] Dr. B.V.A.Rao. "Condition monitoring & cost models for CBM", **Proceedings of the Regional Conference on Recent Trends in Maintenance Management,** 2005**,** pp. 3-10

[4] Oscar Fernandez, Asharf W labib, RalpWalmesly, David J.Petty. "A decision support maintenance management system: Development and implementation", **International Journal of Quality & Reliability Management**, 20(8), 2003, pp. 965-979

[5] Zulkifli Tahir, Anton Satira Prabuwono and Burhanuddin Mohd Aboobaider, "Maintenance Decision support system in small and medium industries: An approach to new optimization model", **IJCSNS international Journal of Computer science and Network security**, l8 (11), 2008, pp. 155-162

[6] Moubray, John. **Reliability-Centered Maintenance**: Industrial Press Inc, 1997.

[7] Andrew .K.S. Jardine, **Maintenance Replacement and Reliability Theory and Applications,** CRC Press, 2005.

[8] Salih O. Duffuaa, A. Raouf, John D. Campbell. **Planning and control of maintenance systems: modeling and analysis**, John Wiley & Sons Publishers, 1999.

[9] HarisKadir Mohammed, K.P.Ramachandran and Rajesh Mangal. Mathematical Models for Optimal Inspection & Preventive Maintenance of Plant Equipment: Some Case Studies, **Regional Conference on Recent Trends in Maintenance Management**, Oman, 2005.

[10] Ashraf W Labib, A decision analysis model for maintenance policy selection using CMMS, **Journal of quality in maintenance engineering**, 10(3), 2004, pp. 191-202

Innovation Promoted by Meta-Engineering
- Mining-Exploring-Converging-Implementing Process -

Hiroshi Suzuki
GE Energy
Akasaka 5-2-20, Minato, Tokyo, JAPAN 107-6111

Yuji Okita
Kanazawa Institute of Technology
Jingumae 1-15-13, Shibuya, Tokyo, JAPAN 150-0001

ABSTRUCT

The authors propose a new creative concept of "meta-engineering" as a dynamic engineering approach that is effective for breakthrough innovation.

A science and technology driven country is expected to play a leading role in addressing increasingly more diversified global issues and in solving them with maintaining harmony of sustainable growth of human society and preservation of environment in recent years. A key for performing such a role lies in innovation by continual orchestration of global issues, science and technology, and new social value.

Addressing merely emerging issues with science and technology is insufficient for realizing breakthrough innovation. The proposal maintains that the spiral process is a key driving force for innovation and solution of challenges that local economy, regional communities, whole countries, and the world are faced with. The processes include "mining invisible and potential/latent issues from bird's eye point of view (M)", "exploring and strengthening necessary science and technologies (E)", "converging these science and technologies to generate solutions (C) " and "implementing them for creating social value (I)". These processes are represented as MECI process.

"Meta-engineering" that the authors propose is different from the engineering approach of today that focuses only on converging of technologies. "Meta-engineering" is to promote radical innovation going down to the root of a superficial problem.

For further deepening and perfection of "meta-engineering," more accelerated investigation and specific demonstration and implementation in parallel are indispensable. Such actions require well-coordinated and intense academia-industry-government collaborations with close interactions with the general public, under the national and international science, technology and innovation policies execution.

The authors emphatically propose to put a high priority on the development of "meta-engineering" as a new approach for innovation and sustainability for generations to come.

Keywords: Innovation, Science and technology, Converging technology, Meta-engineering, Issues.

1. INTRODUCTION

The authors are working as members of the Engineering Academy of Japan (EAJ). EAJ has the Committee on Technology Policy that proposes effective policy of science and technology that are necessary from the standpoint of engineering for society.

Globally cloud computing, smart grid, and iPhone/iPad are coming out as innovation. On the other hand, not so many innovations are coming out from Japan while Japan is said to be extremely good at engineering, capable of making excellent products, and has competent craftsmen. This is the start point of the investigation.

Engineering has many definitions and it is often defined "to design under constraints [1]" in National Academy of Engineering in the US. Or "to provide an optimal solution within a limited given condition." The authors question whether these definitions are sufficient. Totally different answers may be obtained that may lead to innovation, by removing the constraints, instead of by narrowing down. In studying innovation, the authors first took notice of "converging technology," which is an idea developed in the US [2] and the EU[3]. If this idea fits well in Japan, just need to be developed into a Japanese-style converging technology. After analyzing these

activities the authors reached a conclusion that meta-engineering is the key factor for innovation.

2. CONVERGING TECHNOLOGIES IN EUROPE AND US

Converging technologies (CTs) start with identifying what the future challenges will be and exploring what sciences and technologies will be needed for them. In the final proposal of CTs, the four fields of NBIC – nanotechnology, biotechnology, informatics technology, and cognitive science – are identified as the core technologies [4]. It also says that any single field of them is not enough to address global issues and that converging multiple fields are necessary. "Converge" means "to bring together." While the four fields of NBIC are originally independent, they should be converged keeping the original fields. Some fields may merge into a new field if a new field emerges, but the respective original fields must remain as well.

3. CONVERGING TECHNOLOGIES IN JAPAN

In the US where innovation continues, the people are superior in mining potential issues. They find issues to address solutions, and then make all efforts for the solutions. Japanese are good at finding a solution for a given issue under limited conditions, but are rather weak with open questions with no limits nor conditions. Innovation is not realized unless approaching the unseen issues and seeks solutions. It is necessary to think what is behind the visible issue, what the real issues are, and what the hidden issues are.

When a Japanese company does business, it tends to think, "We are capable of doing this. How could we make this into business?" In US Company, however, it has a clear vision "we want to do this kind of business." A project starts in a top-down style, where the top people think what they have, what they don't have, and what they should do. In Japan, the bottom-up style is very strong, where the technology that the company possesses is molded into a new product.

Two typical Japanese products are as examples. First one is an air conditioner. It is highly efficient. It uses intelligent inverters and heat pumps, and utilizes very fine technology. Secondary the hybrid vehicles combine the gasoline internal combustion engine and the battery motor in a sophisticated manner. Two of these are both very excellent in a manner that these solve the visible problems under the given constraints. Since the Japanese are capable of such skills, they try to solve problems in that manner.

4. META-ENGINEERING

The authors decided to name the effort of mining potential issues and solving them by removing the limitations as "meta-engineering." (Fig.1) The definition of Meta-Engineering is as follows:
First we find the hidden and not apparent issues in our society. Then we look for the appropriate science and technologies to solve this issue. In many cases it is difficult to solve the issue by single science and technology and we need to converge several science

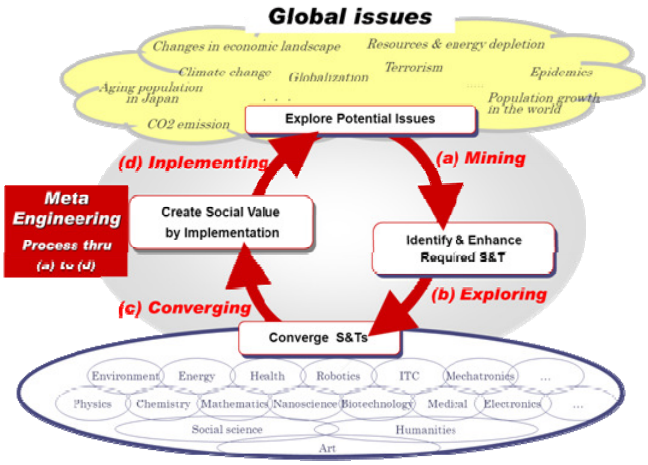


Fig. 1 Meta-Engineering

and technologies. Finally we implement this solution to the real world and get new value to the society.
Several references show the term meta-engineering has been used in some areas [5], [6]. The authors selected this term for this proposal. After considering other candidate names such as "holonic engineering," "comprehensive engineering," "ecological engineering," "transformative engineering," or "Japanese converging technology," etc.. The authors, however, determined as meta-engineering since this name produces an image of metaphysical engineering as a level above current engineering. Meta-engineering places "why" question at the beginning. For instance "why is innovation is necessary in energy field?" One of the answers will be "to keep the global environment and natural resources for our next generation."

Always keeping the above "why" in mind, meta-engineering is to circulate the four processes into a spiral. It begins with mining process to find an invisible and potential or buried issue (M), and then exploring necessary science and technology to solve it

(E). If the issue cannot be solved by existing science and technology, some technologies are converged (C). Finally, the solution to the issue is implemented (I). These processes are represented as "MECI." Another new issue emerges in this process. The image of the four processes turning round and round is important.

One of the reasons of returning to the process of mining a new issue is that innovation is meaningless unless it continues. And so the four processes repeat cyclically.

In that sense, it is a spiral rather than a cyclical feedback. It means that, the world may change by introducing new things, but some other potential issue arises because of that new introduction.

This MECI process is cultivated in meta-engineering field or "Ba" in Japanese.

5. IMPORTANCE OF 'WHY' PROCESS

The most difficult part is mining potential issues. The authors have no specific plan at the moment, but marketing shows some hint.

A salesperson visits a client, and the client says, "I want to drink some juice." (Fig.2) In a Japanese company, the salesperson will purchase a high technology juicer and some fresh fruits, make juice, and take it to the client. The client will be 100% satisfied and may buy the glass of juice for 10 dollars. Though the actual cost maybe 9 dollars to buy the juicer and the fresh fruits. Selling the glass of juice for 10 dollars generates 1-dollar profit. In a Japanese company, this is highly evaluated because the customer satisfaction is 100%.

On the other hand a US company looks at the root of the issue. If the client says, "I want some juice," the salesperson will ask, "Why do you want juice?" (Fig.3) When the client answers, "Because I'm thirsty," the salesperson will sell water. This will solve the client's thirst. Another client may say, "I want some cola." He may be just thirsty. Then, the salesperson will sell a glass of water for 1 dollar to 10 thirsty clients. If he sells water for 1 dollar a glass, the sales will be 10 dollars. Assuming the total cost of 10 glasses of water is 5 dollars, the profit is 5 dollars.

In this example a Japanese salesperson misses the process to identify "what is really necessary." The authors consider this process is important and Japanese should train themselves to acquire this point of view. The Japanese are keen on "how" and "how-

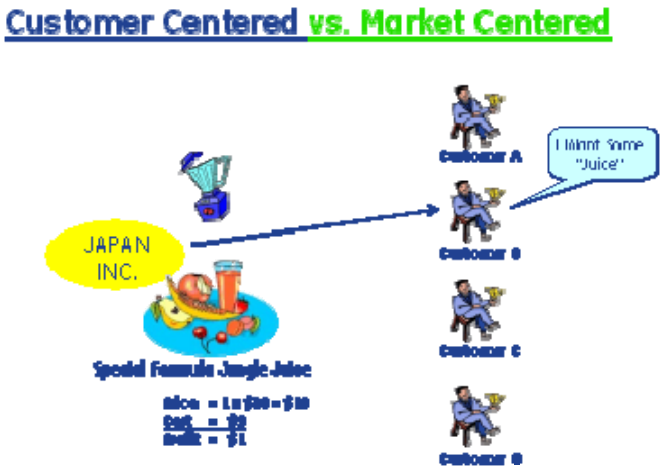


Fig. 2 Customer Centered Approach

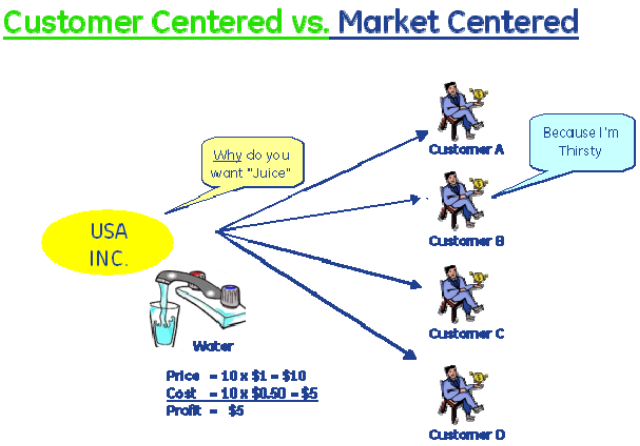


Fig. 3 Market Centered Approach

to." But they need to go beyond into "what is really important" and "why" so that they would mine invisible potential issues. In the example "what" is to know that a client wants juice, "how" is what kind of juice to sell, and "why" is that the client is thirsty.

In conventional engineering, "what" is given as a problem to be solved, and an engineer figures out "how" to make something. It is suggested that an engineer go down to "what" and "why."

Through the discussions in EAJ the authors noticed two major points. One is education: How to help people realize the importance of "why" process. The other is further research of meta-engineering itself. These are the current tasks that the authors propel.

As to education, debating is a common and effective measure in the US. In the course of debating, the settings and perspectives will change. Through debate people will learn switching perspectives.

There was a symposium on privacy and security in the information society, organized by the engineering academies of Japan and the UK. Japan discussed about how pattern recognition could be performed by video cameras and at what angles the cameras should be set. The UK started a discussion on the institutional approach of security conservation, of how to privacy while maintaining national security.

Japanese engineers are accustomed to saying nothing about the system. They focus on technological aspects because they consider that it is what they are expected to do.

Switching perspectives are important for engineers who tend to get fixed a perspective if they stay in one place too long.

US company respects carrier mobility. Someone in sales division may go to marketing division, or M&A one, which the authors regard business development, or do project management. People can experience different types of work to enhance their own expertise. If one stays in a position for 18 months, he/she earns the right to move to another section in some US company. There exist incentives to encourage mobility.

6. MANAGEMENT OF TEHNOLOGY AND META-ENGINEERING

There is a close relationship between management of technology (MOT) and meta-engineering. Management of manufacturing technology, as an example, shows the similarity between MOT and meta-engineering. Here the important thing is not only manufacturing process but also the kinds of product to market, develop, manufacture and sell. Namely the multiplication of "what" and "how" must be considered. In short MOT is the multiplication of "technology" and "management" under the uninterrupted dialogue with market place. Even if you have a good technology, it will be useless without good management. And good management is meaningless without a good technology. Balancing of multiplication of technology and management is important. Meta-engineering can play a significant role here.

There was a book written by Lester and Piore, "Innovation: The Missing Dimension [7]." It says, "Innovation will take place interpretively rather than analytically." The authors felt that this perspective was new to Japan. Therefore it is important to expand engineering to interpretation rather than to analysis only. People with engineering expertise have the knowledge of analysis. So if they enter the interpretive process they may be able to attain meta-engineering.

This meta-engineering is practically realized in the innovation of smart grid in energy area, cloud computing in information area, and will be applied to create innovation in mobility of human beings transportation in future.

7. METHODOLOGY AND APPLIED EXAMPLES

The outline of the methodology of the meta-engineering is shown in Fig. 4.

When one applies meta-engineering to an issue to be studied, the first step to be followed is to state "Why it is the issue to be studied" based on the field on which "What is the issue to be studied." Then it is necessary to dig the "What." The process will become the one shown in Fig. 5.

What can become "why"? Anything will be allowable only if engineering will be workable and is ethically acceptable for the proposition. It is unsuitable if the motion is based on the greed to get money by cheating socially vulnerable people, such as a retired handicapped person or a young child even if it will be easily attained by applying modern ICT technologies.

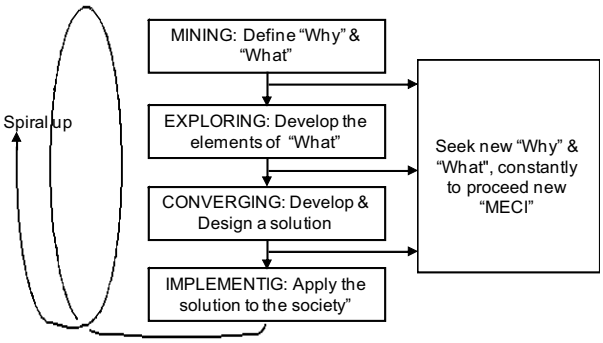


Fig. 4. Outline of the methodology

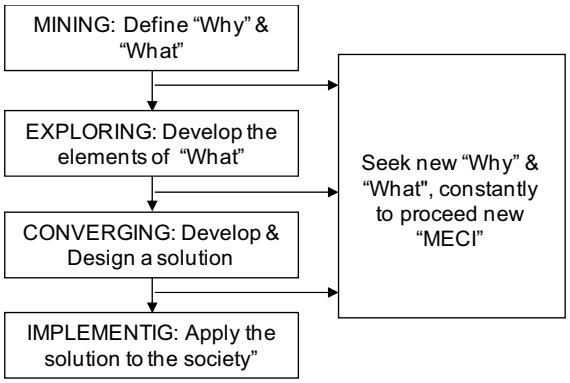


Fig. 5. Mining Stage

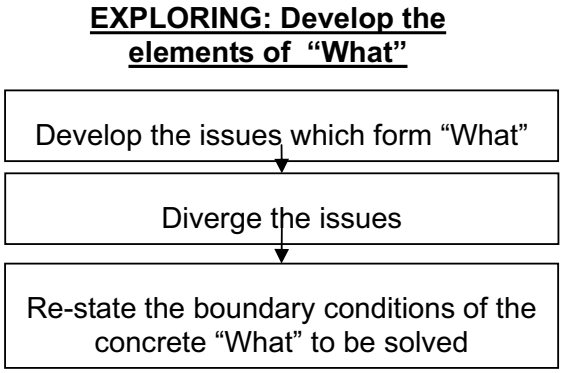


Fig. 6. Exploring stage

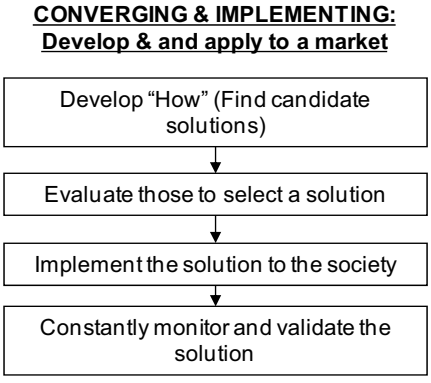


Fig. 7. Converging stage

Let us show an example. Suppose the basic field be an energy problem. Why is the energy problem to be studied? Because it is the science and technology heavily related issue and it has to be overcome for the survival of human beings and for the sustainability of the earth. These are the definitions of "Why."

When one applies meta-engineering to an issue to be studied, the first step to be followed is to state "Why it is the issue to be studied" based on the field on which "What is the issue to be studied." Then digging the "What." The process will become the one shown in Fig. 5.

As far as the proportion of the electric energy continuously increasing to occupy major part in the total energy consumption, electric energy system should be innovated dramatically. More concrete expression of the subject is described as follows. The generation and consumption of electric energy are so rapidly changing that the grid system to connect these should be innovated dramatically. The innovation could be one of the vital solutions of the previously stated "What." It is seldom understood by the general public why the grid plays very important role in the electric power system to use electric energy effectively.

The emerging issue about the power grid is "smart grid." Therefore, the next "What" will be as follows. "Up to today, the efficiency and reliability of the electric power system has been raised by enlarging its scale. Without killing the strong point of the large scale power network, and with promoting electric power consumer be also a small scale power producer, we would have the power system of next generation. To develop the innovative way to go together of large and small is the key issue." This is the meta-engineering "What" statement on the smart grid and above description is the "Mining" stage.

Next is "Exploring." The process shown in Fig. 6 will be followed.

It is necessary to develop the concrete parts, another word, constituents which form "What" of the smart grid. There are many; for instance, solar and wind power generation equipment, electric energy storage including battery for automobile, various electrical household appliances ,and ICT equipment and software to connect these parts. Some of these are in a so called "Green House," others are in the grid. Various types of infrastructures which support metropolitan and urban life would be taken into account. It is sometimes effective to defuse the

analysis intentionally ,so that interesting new issues may be found.

These are all electricity related parts. Here, the distribution and consumption of electricity is connectively considered. Electric current form and voltage height are the issues. Which are the more convenient form, alternative or direct current for a specific purpose? How high voltage is effective for distribution and consumption. The electrical current from solar power panel is at first dc. Electricity from wind power is at first unstable ac. TV and personal computer use very low dc voltage to drive their electronic circuits. Air conditioner and refrigerator receive 50/60Hz ac, then change its frequency, another words, change received 50/60Hz ac to dc then dc to ac of adequate frequency.

Electric energy storage is another important issue. Until today, although pumping up hydro power station has worked effectively as energy storage, and fuel cell storage has continuously developed, the amount of storage has still been insufficient form the smoothed electric load curve point of view. The large number of very small distributed electric power storage, that is an automobile battery, is dramatically changing the situation by being connected to power grid. Together with solar cell, electricity is becoming consumed at the very close place where it is generated. This type of generation-consumption ratio will increase heavily. Then is everything distributed?

The authors think "No." The development of technology and business model by which many types of concentrated and distributed power generation are smartly connected, monitored and controlled would be the winner of the coming electric energy field.

Many other issues are to be considered in this field, such as compatibility, standardization, construction, maintenance, disposal, ownership of equipment and so on. Standardization, maintenance and disposal are especially important to the distributed equipment. Ownership innovation might ease the problem.

It is then important to clarify the boundary condition of these issues, another word to specify the conditions so that each problem would be solvable, by exploring these issues without sticking in a single science and engineering domain. After that, the process to seek candidate solutions by combining the knowledge of related fields or by deepen a specific field is proceeded. Then the most adapted solution is selected (sometime plural solutions are selected), developed, designed and implemented (Fig. 7).

At the final stage, the business incentives are quite important. Not only business model but also social systems such as taxation and regulation (de-regulation) must be innovated.

8. CONCLUSIONS

Meta-engineering can be applied globally for promoting innovation creation. As Japan boasts manufacturing, it should maximize the accumulated experiences as its strength. The MECI process shall be much more activated to create innovation. In order to promote this concept one important activity is education on meta-engineering in engineering course. And the other is to cultivate "Ba" or meta-engineering field on which the MECI process is smoothly carried out.

ACKNOWLEDGEMENTS

This research is currently carried out in a task force of the EAJ and sponsored by Nissan Global Foundation. The members of this work are Yuko Ito (MEXT), Ichiro Katsumata (Sonoba (In-Situ) Research Institute), Yasutoshi Komatsu (Watanabe Co. Ltd.), Yoshio Matsumi (ITOCHU Corp.), Takayuki Nagata (National Museum of Nature and Science), Chie Sato (BizTech Inc.).

REFERNCES

[1]National Academy of Engineering, "The Engineer of 2020." http://www.nae.edu/programs/education/activities10374/engineerof2020.aspx

[2]Converging Technologies for Improving Human Performance, NSF（2002）

[3]Converging Technologies - Shaping the Future of European Societies, EC（2004）

[4]Managing Nano-Bio-Info-Cogno Innovations: Converging Technology Society, NSF（2005）EAJ publication, March 2010: http://www.eaj.or.jp/proposal/CT%20meta%20engineering%20proposal%2005252010.pdf

[5]http://metaengineering.org/problem_space.html

[6] Nagib Callaos, "The Essence of Engineering and Meta-Engineering: A Work in Progress" http://www.iiis.org/Nagib-Callaos/Engineering-and-Meta-Engineering/Engineering-and-MetaEngineering.pdf

[7] R. Lester & M. Piore, "Innovation, the missing dimension." Harvard University Press, Oct. 2004

Review of Empirical Researches on Social Network Sites

Yinan Yu and Yan Zhu
Management Science and Engineering Department in School of Economics and Management, Tsinghua University
Beijing, 100084, China
{yuyn.10, zhuyan}@sem.tisnghua.edu.cn

ABSTRACT

Social network sites (SNSs) intrigue the academic and industry for their heir large numbers of users, broad coverage and astonishing increment speed, and attracts more and more attentions of them. This paper reviews empirical researches on SNSs at individual level, and categorizes them into two parts. The first part investigates the factors affect usages of SNSs and the second part examines what kind of effects the SNSs bring to users. After summarizing the main contents and features of the two kinds of existing studies concerning SNSs, we conclude with considerations for future research.

Keywords
Social network sites (SNSs), Facebook, antecedents, use, social capital, privacy

1. INTRODUCTION

Social network sites (SNSs), such as Facebook, MySpace and LinkedIn have attracted hundreds of millions of users, and become part of users' daily life since their introduction. Four different terms for Facebook have arisen in Hitwise's Top 10 search terms, including "Facebook login" and "www.facebook.com", which account for 3.48% of all searches in the US among the top 50 terms. Furthermore, Facebook.com has firstly topped Google.com as the most-visited website of the year in March 2010 [1]. SNSs have been also experiencing incredible growth in China since 2003. Nowadays, many SNSs like renren.com (which began among students in high schools and universities and is a successful imitator of Facebook), kaixin001.com (which aims at white-collar population), douban.com and so on, attract lots of users by online community. At the same time, researchers, mass media and business persons in China have shown a lot of interests on SNSs and their trendy.

There are variety kinds of SNSs, which allow individuals to present themselves, articulate their social networks, and establish or maintain connections with others [2]. Some sites, such as Facebook.com, renren.com, etc., support the maintenance of existing off-line social networks and some others help strangers connect according to their shared interests and similar purposes, like dating site (e.g. Match.com) and business-related site (e.g. Linkedin.com). Some sites cater to diverse audiences, while others attract people based on common language, racial, or religious. SNSs provide a great deal of services, such as creating users' own space, sharing photos, maintaining blogs, instant messaging (IM), and so on.

Scholars from disparate fields have examined SNSs usage in order to understand the practices, implications, cultures, and meanings of the sites, as well as users' engagement with them [3]. Most of them select Facebook.com as their research platform, which is the most famous social network site. We review empirical researches that are directly concerned with SNSs at individual level in this paper, and then categorize them into two parts. The first part investigates which factors affect usages of SNSs, and the second part examines what kind of effects usages of SNSs bring to users. After summarizing the main contents and features of the two kinds of existing researches concerning SNSs, we conclude with considerations to set a stage for future research.

2. BACKGROUND

Definition

According to the definition of Wikipedia, SNS is "an online service, platform, or site that focuses on building and reflecting of social networks or social relations among people" [4]. Boyd and Ellison define SNS as a kind of web-based service that allows individuals to three kinds of things. First, construct a public or semi-public profile within a bounded system. Second, articulate a list of other users with whom they share a connection. Third, view and traverse their list of connections within the system [3]. All the definition of constructors mentioned in this article are shown in Table I.

From the above definitions, we can conclude some common features of SNSs and the main difference distinguishing SNSs from instant message and blogs. That is the nature of socializing, which makes SNSs essentially consist of a representation of each user (often a profile) and his/her social links, and an individual-centered service.

Table I. Definitions of constructers

Constructer	Definition
SNS	a kind of web-based service that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections within the system [3].
self-efficacy	the belief in one's capabilities to organize and execute the courses of action required to produce given attainments [5]
need for cognition	individual's tendency to engage in and enjoy effortful cognitive endeavors [5]
need to belong	people need to be loved and socially accepted collective self-esteem: that aspect of the individuals' self-concept which derives from their knowledge of their membership in a social group together with the value and emotional significance attached to that membership [5]
Collective self-esteem	the aspect of the individuals' self-concept which derives from their knowledge of their membership in a social group together with the value and emotional significance attached to that membership [5]
situation factors	strength and nature of relationships with other members, perceived risk, and privacy and security issues and so on
privacy dilemma	privacy needs and the need for sociability and content sharing [13]
social capital	resources accumulated through the relationships among people [2]
bridging social capital	Bridging social capital is linked to "weak ties", which are loose connections between individuals who may provide useful information or new perspectives for one another but typically not emotional support [2]
bonding social capital	bonding social capital is found between individuals in tightly-knit, emotionally close relationships, such as family and close friends [2]
maintained social capital	the ability to maintain valuable connections as one progresses through life changes [2]

Development

SNS has existed over a decade, and its development mainly contains three phases. The first phase began in 1997, when the first SNS, SixDegrees.com, was launched. It allowed users to do the basic and most essential work, such as creating profiles, listing friends and so on [3]. The second phase started in 2001, when Ryze.com was launched. Some other similar SNSs, such as Linkedin etc, were established during this period, and they focused on a certain topic or function, such as business. The third phase began after the year 2003. Many multimedia-shared SNSs, such as Flickr.com and YouTube.com, arose from the development of UGC (user generated contents). After that, comprehensive SNSs like Facebook started to emerge.

3. ANTECEDENTS OF SNS USAGE

The first kind of researches focuses on which individuals use SNSs, how and why they use them, and what kind of antecedents affect usage of SNSs. As Figure 1 shown, passing researches involve three aspects: self-description, socioeconomic factors and situation factors.

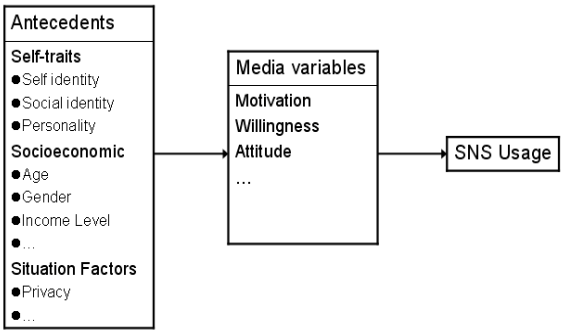


Figure 1. Antecedents of SNS usage

First, in self-trait part, some researchers pay attention to the influence of social identity. Gangadharbatla found that Internet self-efficacy, need to belong, and collective self-esteem have positive influences on attitudes toward SNS use [5]. Furthermore, attitudes fully mediate the relationship between willingness to join SNS and Internet self-efficacy and need to belong and partially mediate the relationship between willingness to join and collective self-esteem. Additional psychological predictors like self-construal [6] are discussed in detail. Others focus on different personality's influence on SNSs usage [7][8][9]. For example, Orr et al. support that "shyness" is positively correlated with the time spent on Facebook and negatively correlated with the number of Facebook "friends" [9].

Second aspect comes to socioeconomic factors. Dimensions like age, gender, income level, culture, and ethnicity are usually discussed [6][10][11][12]. Hargittai found that a person's gender, race and ethnicity, and parental educational background are all associated with use of SNSs [12]. For example, Hispanic students are significantly more likely to use MySpace than are Whites in the sample, while Asian and Asian American students are significantly less likely to use MySpace.

Third, situation factors refer to strength and nature of relationships with other members, perceived risk, and privacy and security issues and so on [5]. Scholars pay much attention to the privacy issues [13][14][15][16]. Brandtzæg et. al study the "privacy dilemma" issue, especially on how multiple social ties and groups on SNSs affect users' social privacy, and what kind of differences in practices and experiences of sociability, content sharing, and privacy are, between younger and older adults [13]. One of their finding, for example, is that younger users are more skilled social networkers and more likely to be aware of and to change privacy settings compared to the older.

This first point of view of research attracts many scholars from different disciplines. Researchers from not only information system and management, but also computer science, psychology, sociology and mass communication show a lot of interests and do much of research work on this fangle, in order to figure out why so many individuals, especially young people, take their time and energy for SNSs.

4. SNS USAGES AND OUTCOMES

As illustrated in Figure 2, second kind of studies focuses on SNS usages and their outcomes.

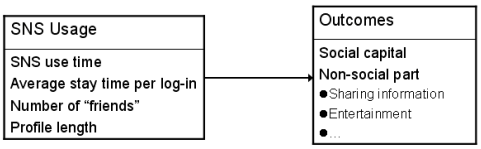


Figure 2 SNS usage and outcomes

Researches of SNSs usages have pay attention to different aspects. Some scholars conclude what SNSs can provide for their users. According to Boyd and Ellison [3], SNSs are "web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections" within the system. Others focus on researching users' behaviors on some special usages or applications of SNSs [17] [18]. For instance, Golder et. al. [18] select two kind of usages of SNS， messaging and poking, and use them as proxies to reflect college students' daily, weekly and seasonal schedules and to discuss their use as a way for college students to support both distant and geographically proximate relationships. Additionally, Lampe et al. [19] look at how use and perception of Facebook.com has changed over time indicated by their three consecutive years of survey data and interviews. They find that there was little change in Facebook use over time in most ways. Finally, SNSs usages are always broken down and measured by constructs like SNS use time, average stay time per log-in, number of "friends" in SNS, and profile length [6][9].

Usages of SNSs bring several positive outcomes. First, SNSs can help maintain existing offline relationships or enhance existing relationships [20][21]. According to Lampe et al., their research objects through both survey and interview show that they are typically using the site to maintain lightweight contact with relationships they have developed offline [19].

Second, scholars examine the relationship between the use of SNSs and the formation and maintenance of social capital (benefits received from individuals in one's network) [2][21]. Ellison et. al. [2] found strong positive associations between usage of SNSs and the three types of social capital (bonding social capital, bridging social capital and maintained social capital), with the strongest relationship being to bridging social capital.

Third, SNSs can satisfy users' non-social needs such as entertainment, recreation, searching information and sharing contents. In addition, Chen started to research what factors would influence the continue use of professional virtual communities (PVC) by using Expectation-Confirmation Theory (ECT), and found that

the contextual factor and technological factors both exert significant impacts on PVC participants' continuance intentions [22].

As a nature of sociability of SNSs, large part of the second kind of research discuss social capital issue, which is accepted as one of the most important benefits of SNSs and the key factors for their continuous development.

5. DIRECTIONS FOR FUTURE RESEARCH

As described in the prior section, SNSs have attracted lot of researchers from disparate disciplines, different region and culture to do research on this topic, because of their broad coverage and amazing increment speed. Given the continuing interest and empirical attention to SNSs, we provide this review paper and some potential directions for future researches.

Usually, large numbers of studies about predictors of the adoption of certain new technology will emerge during the initial stage of its introduction, and so is the same for SNSs. Many researches about antecedents of SNSs usage have been done, and it has important practical implication for the improvement of design of SNSs. As SNSs are complicated systems, there are still lots of influencing factors needed to discuss in the future research, which could suggest devising effective strategies that take into account differing levels of users for website owners, and could be helpful to enhance perceived usefulness and perceived ease of use [23][24] of SNSs for users in order to increase membership and participation, and bring more revenue for operators of the sites. Furthermore, a relatively integrated conception model including different aspects of antecedent variables maybe useful to understand this issue completely.

Other potentially promising avenue for future research could be consequences of SNS adoption. As discussed above, social capital is center topic in this research field, and several studies have discussed the positive impact SNSs usage can bring to form or maintain social capital for users.

However, as the number of friends on SNSs increases and different kinds of friends or groups hold different kinds of relationship with the user, one user usually face diverse social capital in the same time in this transparent environment. Generally, an individual performs differently when facing distinct social capitals. As a result, on one hand, we need to study social capital itself and add some new meanings to it in surrounding of SNSs, on the other hand, we need to dig deeper into what kind of social capital could be mixed and in what way it is suitable in socially transparent environments such as SNSs [13].

Sociability is not the only positive outcome of SNSs, as we mentioned before. However, there is little research on benefits that has little to do with sociability or connections about SNSs. Future research may also exam this "non-social" benefit of SNSs by using theories or models from different disciplines.

Finally, although we divide research on SNSs into two parts, these two parts are not completely divorced and have no relationship with each other. Actually, interrelation may exist between these two parts. For instance, benefits of using SNSs may have effects on the motivation for SNSs usage, and people with common motivations may get different benefits from SNSs usage. That is to say, this direction may be a meaningful research topic in the future.

REFERENCES

[1] eMarketer (2006), "Is Facebook Really No.1?". Available at http://www.eMarketer.com (accessed January 11, 2011).

[2] Nicole B. Ellison, Charles Steinfield, Cliff Lampe (2007). The Benefits of Facebook ''Friends:'' Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication* 12 (2007) 1143–1168

[3] Danah M. Boyd, Nicole B. Ellison (2008). Social Network Sites: Definition, History,and Scholarship. *Journal of Computer-Mediated Communication* 13 (2008) 210–230

[4] Wikipedia. 2011. Social Network Service. Retrieved April 1, 2011 from World Wide Web: http://en.wikipedia.org/wiki/Social_network_service

[5] Harsha Gangadharbatla (2007). Facebook Me: Collective Self-Esteem, Need to Belong, and Internet Self-Efficacy as Predictors of the iGeneration's Attitudes toward Social Networking Sites. *Journal of Interactive Advertising,* Special Issue on Online User Generated Content June 9, 2007

[6] Jang Hyun Kima，Min-Sun Kima, Yoonjae Nam (2010). An Analysis of Self-Construals, Motivations, Facebook Use, and User Satisfaction. *International Journal of Human-Computer Interaction* 6(11–12) 1077–1099, 2010

[7] Emma L. Pelling, Katherine M. White (2009). The Theory of Planned Behavior Applied to Young People's Use of Social Networking Web Sites. *CyberPsychology & Behavior* Volume 12, Number 6, 2009

[8] Michael A. Stefanone, Chyng-Yang Jang (2008). Writing for Friends and Family: The Interpersonal Nature of Blogs. *Journal of Computer-Mediated Communication* 13 (2008) 123–140

[9] Emily S. Orr, et.al. (2009). The Influence of Shyness on the Use of Facebook in an Undergraduate Sample. *CyberPsychology & Behavior* Volume 12, Number 3, 2009

[10] Mariea Grubbs Hoy, George Milne (2010). Gender Differences in Privacy-related Measures For Young Adult Facebook Users. *Journal of Interactive Advertising*, Vol 10 No 2 (Spring 2010), pp. 28 - 45

[11] Yong Gu Ji, et. al (2010). The Influence of Cultural Differences on the Use of Social Network Services and the Formation of Social Capital. *International Journal of Human-Computer Interaction*, 26(11–12), 1100–1121, 2010

[12] Eszter Hargittai (2008). Whose space? Differences Among Users and Non-Users of Social Network Sites. *Journal of Computer-Mediated Communication* 13 (2008) 276–297

[13] Petter Bae Brandtzæg, Marika Lüders and Jan Håvard Skjetne (2010). Too Many Facebook "Friends"? Content Sharing and Sociability Versus the Need for Privacy in Social Network Sites. *International Journal of Human-Computer Interaction*, 26(11–12), 1006–1030, 2010

[14] Dianne M. Timm, Carolyn J. Duven (2008). Privacy and Social Networking Sites. *New Directions for Student Services* No. 124, Winter 2008

[15] Sonia Livingstone (2008). Taking Risky Opportunities in Youthful Content Creation: Teenagers' Use of Social Networking Sites for Intimacy, Privacy and Self-expression. *New media & society*, 10 (3). pp. 393-411. ISSN 1461-4448

[16] Ralph Gross, Alessandro Acquisti (2005). Information Revelation and Privacy in Online Social Networks (The Facebook case). *WPES'05,* November 7, 2005, Alexandria, Virginia, USA.

[17] Minsun Shim, Min Ju Lee, Sang Hee Park (2008). Photograph Use on Social Network Sites among South Korean College Students: the Role of Public and Private Self-Consciousness. *CyberPsychology & Behavior* Volum 11, Number 4, 2008

[18] Golder, S., Wilkinson, D. and Huberman, B.A. (2007), Rhythms of Social Interaction: Messaging within a Massive Online Network. in *3rd International Conference on Communities and Technologies (CT2007).* (East Lansing, MI, 2007), Springer.

[19] Lampe. C., Ellison, N. B., & Steinfield. C. (2008, November). Changes in use and perception of Facebook. CSCW '08, San Diego, California

[20] Kenski, K., & Stroud, N. J. (2006). Connections between Internet use and political efficacy, knowledge, and participation. *Journal of Broadcasting & Electronic Media*, 50, 173–192.

[21] Sebasti ´an Valenzuela, Namsu Park, Kerk F. Kee (2009). Is There Social Capital in a Social Network Site?: Facebook Use and College Students' Life Satisfaction, Trust, and Participation. *Journal of Computer-Mediated Communication* 14 (2009) 875–901

[22] Irene Y. L. Chen (2007). The factors influencing members' continuance intentions in professional virtual communities – a longitudinal study *Journal of Information Science*, 33 (4) 2007, pp. 451–467

[23] Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly, 13*, 319–340.

[24] Venkatesh. V. Morris, M. G, Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*, 425–478.

Integrating Professional Software in Undergraduate Civil Engineering Education

Suleiman ASHUR, Ph.D., P.E.
Mohammad ALHASSAN, Ph.D.
Dong CHEN, Ph.D., P.E.
Civil Engineering Program-Department of Engineering, Indiana University-Purdue University Fort Wayne
Fort Wayne, IN 46805, USA

and

Shan GUNAWARDENA, P.E., PTOE
Traffic Engineering, City of Fort Wayne
Fort Wayne, Indiana 46802

ABSTRACT

The effective use of computing in undergraduate civil engineering education is becoming an integral part of education in the US institutions. This paper presents the use of professional software in the following three areas of civil engineering: Hydraulics Engineering, Structural Analysis and Design, and Transportation Engineering. In addition, the paper will include an example on the use of AutoCAD in real life project during the freshmen year.

The paper will present a brief description of software, how it is utilized including examples from individual classes and capstone course. The following software will be presented:

a) CAD Software: AutoCAD 2010/2011.
b) Construction Management package: Microsoft Project.
c) Structural Analysis and Design packages: SAP2000 and ETABS
d) Transportation packages: The Synchro and SimTraffic,
e) Hydraulics package: KYPipe 2010.

In addition, the use of Excel in construction management and transportation classes will be introduced. A discussion on the pros and cons of using these packages as well as a summary of students' comments and evaluation of these packages will be presented.

Keywords: Professional Civil Engineering Software, freshmen and Capstone courses, Construction Management, Structural Analysis, Transportation, and Hydraulics.

1. INTRODUCTION

Using technology in education is becoming an integral part of civil engineering (CE) education in the US. Computer software packages can be used in many tasks including but not limited to collecting, storing, analyzing, and displaying data[1]. A study conducted by the American Society of Civil Engineers' Task Committee on Computing Education of the Technical Council on Computing and Information Technology found that the use of Computer-aided Design (CAD) and Spreadsheet software ranked the highest among educator and practitioners as shown in Table 1[2]. A list of all software packages are presented in the reference 2. In the same study, practitioners prefer to have graduates with solid theory and fundamentals, be able to do hand calculations, and to have real engineering experience before graduation. N.S. Grigg, et al. reported that the CE profession will be left behind if it ignore technologies and does not take advantage of it[3].

Table 1: Selected Skills from the "Survey of Importance Results"[2]

Computer Skill	Practitioners		Educators	
	Rank	Rating[(a)]	Rank	Rating[(a)]
Computer-aided Design	3	4.0	3	3.9
Construction Software	10	3.0	11	3.2
Spreadsheet	1	4.5	1	4.4
Structural software	6	3.3	6	3.6
Transportation software	12	2.9	7	3.5
Water Resources	8	3.1	9	3.4

[(a)] 5: most important, 1 = least important

One of the major challenges for engineering educators is that technology and practice of engineering increases in its complexity with time. Accordingly, there will never be enough time to cover all of the required topics in depth[4]. However, it is important for educators to remain current with ongoing trends in their field in order to present a general overview of technologies that are used in the workplace. Numerous studies have been conducted to develop new technologies and methods that can be introduced to students in order to allow them to become successful engineers upon exiting universities. In a study conducted by Romero and Museros suggested that computers could play an important in engineering education and in particular in structural engineering education in a variety of ways, one of which is using commercial design programs[5].

2. PROFESSIONAL SOFTWARE IN CIVIL ENGINEERING EDUCATION AT IPFW

The new civil engineering program at Indiana University-Purdue University Fort Wayne (IPFW) was established in 2006.

One way to provide all students with real life experience is by requiring a real life project in most of the engineering courses. In order to achieve this goal, students have to do the calculations for the project by hand and then use appropriate software in their project. The flowing is a summary of selected software used in several areas of civil engineering courses at IPFW including Graphic Communication, Construction Engineering and Management, Structural Engineering, Transportation Engineering, and Hydraulics courses.

The new civil engineering program at Indiana University-Purdue University Fort Wayne (IPFW) was established in 2006. One way to provide all students with real life experience is by requiring a real life project in most of the engineering courses. In order to achieve this goal, students have to do the calculations for the project by hand and then use appropriate software in their project. The flowing is a summary of selected software used in several areas of civil engineering courses at IPFW including Graphic Communication, Construction Engineering and Management, Structural Engineering, Transportation Engineering, and Hydraulics courses.

Computer Aided Design Software

Student at IPFW learn AutoCAD software in their freshmen year in Graphic Communication and Spatial Analysis course. It is a two-hour course that teaches student how to communicate engineering concepts effectively using AutoCAD, writing memos, and conducting presentation. Students are required, in addition to assignments and exercises, to do two projects. One the project is to a layout for a building and design a parking lot. The goal of this project is to expose students early to some aspects of professional life. The proposed lot and layout of the new building are shown in Figure 1.

Figure 1 (a): Proposed lot for new building
(Source: Google Maps accessed on January 21, 2009)

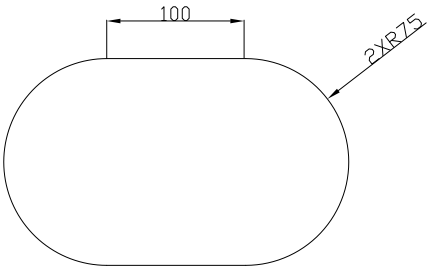


Figure 1 (b): New Building layout, dimensions are in feet.

Students are required to draw the lot to scale with access points to avoid congestions on the street, to locate and draw the building on the lot considering maximizing the utilization of excess land of the lot for the parking spaces, and to design and draw the layout for the parking lot around the building. They shall comply with handicap acceptability act with sufficient parking stalls allocated for handicapped parking. Student needs to communicate the information requested in the project in a brief, 1-2 page memo.

Students have difficulty at the beginning in drawing the lot to scale. They frequently ask about the right number of the parking spaces or what is the best location for the building. Although the project statement provide them will specifics, it is difficult for them at the beginning to handle the new concept of multiple solutions. However, the students' response to this project is mostly positive. Figure 2 show sample of student work.

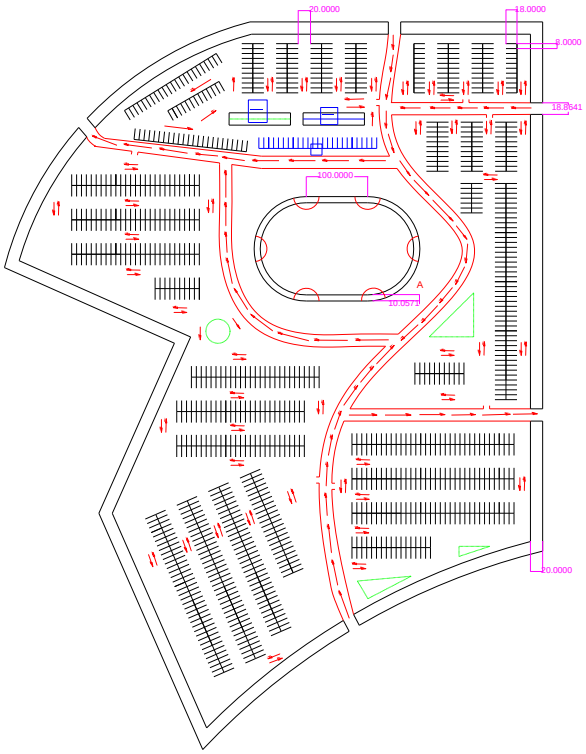


Figure 2: Sample Solution of the Parking Lot Project

Construction Management Software

Excel and Microsoft Project are used heavily in the construction management course. The project in this course requires students to work in a team in a virtual contracting company, and bid on a project. Students are provided with the bidding documents of the project including drawings, specifications, geotechnical report, etc. The goal of this project is to have students acquire a professional life experience by reading blue prints (drawings) and other contract documents. Acting as a contractor, students are required to prepare quantity takeoff, cost estimate for the project, and prepare a detailed schedule of the project activities using Microsoft Project software. Sample of student work in Excel and Project are shown in Figure 3 and 4. Students' feedback on the use of software and the project are positive. However, they would like to learn Excel early on in their life. Usually, students learn fundamentals of Excel in their freshmen year. This course requires high level of skills in Excel which introduced to them during the lab.

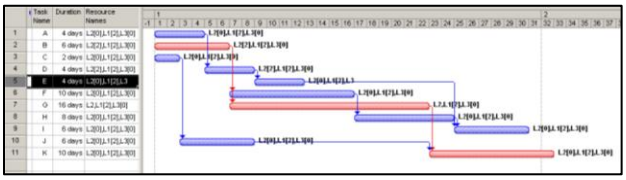


Figure 3: Resource Allocations using Microsoft Project.

Quantities Take off for Foundations

Structural Element (page)	Rebar # (H or V)	Rebar Length (ft)	λ	Rebars / ft	Length (ft)	Quantity	Total Length (ft)	Weight (lb/ft)	Total Weight (lb)	Nominal Diameter (in)
Foundation Wall 4 / S3.1	# 4 H	577.833	1.33	1.5	16.17	23	13434.6	0.668	8974.3	0.5
	# 8 V	17.167	1	1	577.83	579	9936.6	2.670	26530.8	1
	# 4 V	16.167	1	1	577.83	579	9357.8	0.668	6251.0	0.5
Foundation Wall Footings 4 / S3.1	# 5 H	577.833	1	2	1.00	2	1155.7	1.043	1205.4	0.625
	# 4 V	1.917	1	2	577.83	1157	2216.9	0.668	1480.9	0.5
Exterior Wall 15A / S3.2-3	# 4 H	69.290	1.33	1.5	4.00	7	485.0	0.668	324.0	0.5
	# 4 VR	5	2	0.5	69.29	35.645	178.2	0.668	119.1	0.5
	# 4 VL	4	2	0.5	69.29	35.645	142.6	0.668	95.2	0.5
Exterior Wall Footings 15A / S3.3-4	# 5 H	69.290	1	2	1.00	2	138.6	1.043	144.5	0.625
	# 4 V	1.750	2	1	69.29	69	121.3	0.668	81.0	0.5
Piers P1 / S2.1	# 8	13.667	-	-	-	8	109.3	2.670	291.9	1
Pier footer F8 / S2.1	#6	6	-	-	-	12	72.0	1.502	108.1	0.75

where,

- Rebar # (H or V) = Reinforcing Steel Bar Number (Horizontal or Vertical and Left or Right if applicable)
- λ = Frequency of Steel Rebars (1 rebar every *x* feet)
- Rebars / ft = Number of Rebars per Linear Foot
- Length (ft) = Length of the Structural Element
- Weight (lb/ft) = Weight values for each rebar. Source: en.Wikipedia.org/wiki/rebar

Figure 4 Sample of Part of Steel Calculations Using Excel.

Structural Analysis Software

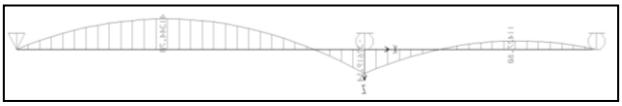
SAP2000 and ETABS are used at IPFW to help the civil engineering students gain real world experience with some of the most powerful software packages available to professionals. Students are exposed to the software step by step of modeling, analysis, and design in a sequential approach through successive core and elective courses including Structural Analysis, Reinforced Concrete Design, and Steel Design. SAP2000 is intended for use on structures such as bridges, dams, stadiums, industrial structures, and buildings. ETABS is a special version of SAP2000 that is used mainly for buildings[6].

Figures 5 show the results of SAP2000 for the shear and moment diagrams as well as the deflection curve for the two-span continuous beam. Figure shows the work of students in a senior design project to design, model and analyze a pedestrian bridge. The deformed view of bridge with wind load applied is shown in Figure 7.

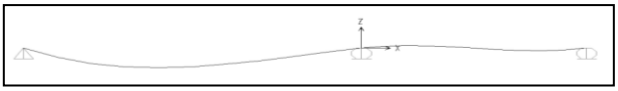
Transportation software

Excel, Highway Capacity Software, Synchro, and TrafficSim are used in the Transportation courses including Transportation Engineering, Traffic Engineering, Transportation Planning and Senior Design. In the Transportation Engineering course, one of the project statements required the use of Excel in developing a program for highway geometric design; from simple curves to compound curves; both vertical and horizontal. This project was linked to ABET outcomes that address the ability of students to design a system, component, or process to meet desired needs within realistic constraints and, measure student's ability to use the techniques, skills, and modern engineering tools necessary for engineering practice. Figure 8 present the input screen for the design of simple and/or compound curve.

(a) Shear diagram for continuous beam



(b) Moment diagram for continuous beam.



(c) Deflection curve for continuous beam.

Figure 5: Structural Analysis of continuous beam using SAP2000.

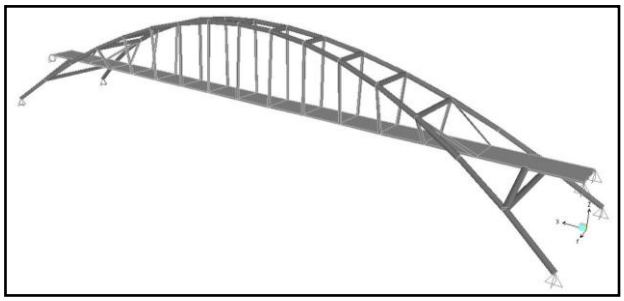


Figure 6: Extruded view of pedestrian bridge.

Synchro plus and SimTraffic are one of the most commonly used software packages in traffic engineering. Some of the important features of Synchro are the analysis tool for traffic engineers include the use of methodologies developed in the Highway Capacity Manual, the effectiveness in evaluating signals and signal systems in arterial networks, the ability to evaluate multiple intersections with a single controller such as Michigan Lefts, diamond interchanges and closely spaces intersections. Synchro's limitations include its limited analytical capabilities of roundabouts and some problems with extremely skewed intersections.

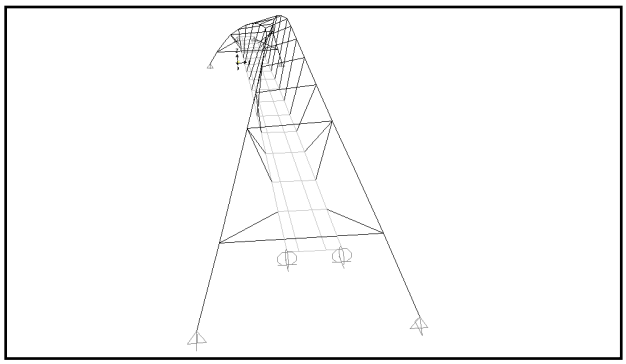


Figure 7: Deformed view of bridge with wind load applied.

Figure 9 shows part of a senior design project entitled "Transportation Network for IPFW." The main objectives of the project is to design a transportation network for the IPFW campus that adequately handles the 60% projected growth of the campus population and facilitates the growth objectives of the Master Plan, while promoting safety. The figure presents the existing design and the proposed design.

INPUT			
Simple Curve enter **1**, Compund Curve enter **2**	CT=	1	
Simple Curve Data			
Multiple of the Stationing (ex. 25, 50, 100)	n =	100	ft
u = Speed [mph]	u =	-	mph
PC = Point of Curvature	PC =	4+65	
Length of Curve 1 (max 68 STA)	L_1 =	183	ft
Radius of First Curve	R_1 =	500	ft
User Deflection Angle	D_1 =	-	degrees
Degree Angle of Simple Curve 1	θ_1 =	34	degrees
Second Curve Data, For Compund Curve			
PCC = Point of Compound Curve	PCC =	6+00	
Length of Curve 2 (max 68 STA)	L_2 =	-	ft
Radius of Second Curve	R_2 =	350	ft
Degree Angle of Simple Curve 2	θ_2 =	26	degrees
User Deflection Angle	D_2 =	-	degrees
Simple Curve Output Data			
Point of Curvature	PC =	4+65	
Radius	R_1 =	500.00	ft
Length (max 68 STA)	L_1 =	183.00	ft
Tangent Length	t_1 =	152.87	ft
External Distance	E_1 =	22.85	ft
Middle Ordinate	M_1 =	21.85	ft
Degree of Curvature	θ_1 =	34.00	degrees
Deflection Angle	D_1 =	11.459	degrees
Angle to determine first full STA	δ_{11} =	6.503	degrees
Angle to determine last STA	δ_{12} =	0.000	degrees

Figure 8: Input and Output Table for Simple Horizontal Curve

Hydraulics Engineering Software

In the Hydraulics Engineering course, students learn sources and distribution of water in urban environment, including surface reservoir requirements, utilization of groundwater, and distribution systems; analysis of sewer systems and drainage courses for the disposal of both wastewater and storm water;

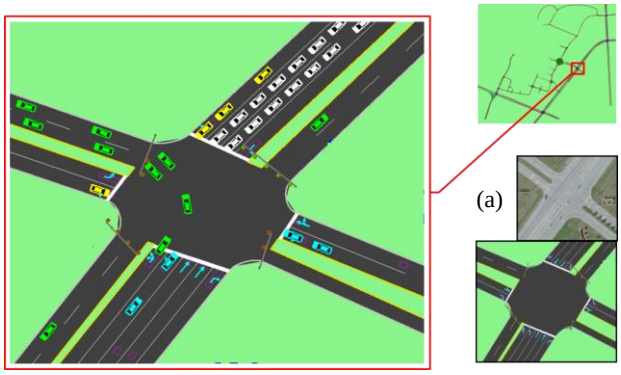


Figure 9: Input and Output Table for Simple Horizontal Curve ([a]) Source: Google Maps accessed on February 21, 2008

pumps and lift stations; and urban planning and storm drainage practice. All these contents are based on the scientific and engineering principles of conservation of mass, energy, momentum, and force balances. The specific laws and equations include Manning's law, Darcy's law, Bernoulli equation, Hazen-Williams equation, and Darcy–Weisbach equation. The students are required to understand and be familiar with all these fundamental principles, which are incorporated in all of the exams, exercises, homework, and projects.

One of the course learning outcome is "to identify and use modern computer software to analyze and design different water and wastewater systems" to the course. This outcome was added because software modeling is a common tool and important platform for current engineers to design and simulate different hydraulic applications.

KYPipe is professional software for hydraulic analyses. The newest version Pipe2010 has a common user interface linked all of the calculation models including hydraulic, water hammer, steady state, and transient systems. However, all kinds of hydraulic software, including KYPipe are just tools. They cannot replace scientific and engineering principles. Students must understand the theories and mechanisms of hydraulics behind the software. As a result, students can be trained as masters with a strong scientific and engineering background, instead of software operators merely. Accordingly, the fundamental principles and hand calculations are arranged to be taught ahead of the software modeling. Afterwards, software is taught using simple examples followed by a couple of design projects to reinforce the skills of computer modeling learnt by the students.

The interface of KYPipe is pretty user-friendly. Generally, left-click of mouse makes nodes and right-click of mouse connects nodes by pipes or conduits. The properties of each node or conduit can be entered or modified in the pull-down window by selecting it. Figure 10 shows an example of a layout pipe network in Pipe2010. In this example, the node of *R-1*, *R-2*, and *R-3* is reservoir, which has a constant water surface elevation and zero velocity. The working curve was given for *Pump-1*. With the assigned pipe materials and diameters, the water surface elevation of each reservoir, and the elevation of each node, the flow rate in every individual pipe was calculated in seconds as shown in Figure 11.

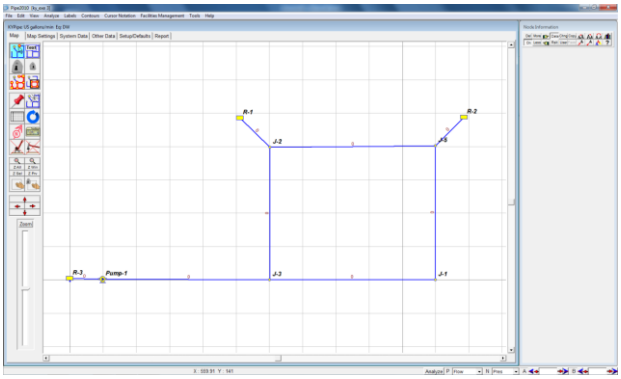


Figure 10: Layout network of a piping system in Pipe2010.

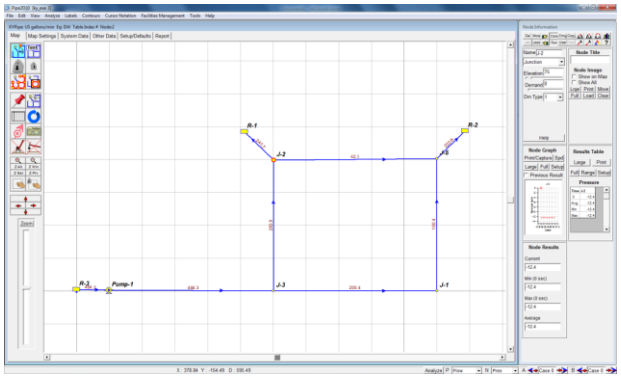


Figure 11: Simulation results showing the flow rate for the piping system by Pipe2010.

3) STUDENT FEEDBACK

The overall students' feedback on using the software is positive. The following is a selected sample of their input:

- Structural Software: the sample of students surveyed used SAP2000/ETABS throughout various stages of their design projects. When students asked on using the software packages in their structural design of their projects, 63% found it very helpful and 37% helpful (37%). In addition, 100% of the students either recommend or strongly recommend using SAP2000 and ETABS in structural engineering courses.

- Excel in Transportation: the results of the assessment indicate that student students have a major improvement of their skills in Excel and in their ability to model and develop engineering design.

- Water Resources: upon completion of the course, a survey was conducted at the end of 2010 to collect comments and suggestions from the students regarding the course and KYPipe software. 60% of the students listed KYPipe as the most favorite part of the course. They were happy to see the software gave them the correct answers within seconds, compared to otherwise difficult, complicated and time-consuming hand calculations. However, although KYPipe has a pretty strong function in simulating pressurized piping system; in storm sewer area, the students found its function and interface could not compete with EPA's Storm Water Management Model.

4) CONCLUSIONS

It is critical for any civil engineering program to focus on the fundamentals of engineering principles, to provide solid theoretical background of the concepts, and expose students to real world projects. Using the software in civil engineering education should be as a tool and in way replace learning sound scientific and engineering principles. However, software can have the advantage of supplementing this knowledge with tool to analyze complicated real world projects. In addition, using these software packages in the Capstone Senior Design project supplies the students with very powerful tools needed not only to complete the project successfully, but also to provide them with a skill that may be attractive to potential employer.

5) BIBLIOGRAPHY

[1] S. Fenves and W. Rasdorf, "Role of ASCE in the advancement of computing in civil engineering," **Journal of Computing in Civil Engineering**, Vol. 15, No. 4, 2001, pp. 239–247.

[2] O. Abudayyeh, H. Cai, S. Fenves, K. Law, R. O'Neill, and W Rasdorf, "Assessment of the Computing Component of Civil Engineering Education." **Journal of Computing in Civil Engineering**, Vol 18, No. 4, 2004, pp. 187–195.

[3] N. Grigg, M. Criswell, D. Fontane, and T. Siller, "Information Technology in Civil Engineering Curriculum," **Journal of Professional Issues in Engineering Education and Practice**, Vol. 131, No. 1, 2005, pp. 26–31.

[4] N.S Grigg, et al. "Integrated Civil Engineering Curriculum: Five-Year Review," **Journal of Professional Issues in Engineering Education & Practice** Vol. 130, No. 3, 2004, pp. 160-165.

[5] M.L. Romero and P. Museros, "Structural Analysis Education through Model Experiments and Computer Simulation," **Journal of Professional Issues in Engineering Education & Practice**, Vol. 128, No. 4, 2002 pp. 170.

[6] http://www.csiberkeley.com/company_about.html. Accessed on August 7, 2009.

The Importance of Interdiscipliner Studies in Engineering Education

Deniz GUNEY
Yildiz Technical University, Faculty of Architecture, 34349, Istanbul, Turkey

ABSTRACT: As the world's dependence on technology increases, there is no question that the demand for engineers is increasing. Engineering is a key component of innovation and technological society. Engineering education system is the most important subject for the future of the society. For large scale projects, many different professions such as architects, civil engineers, mechanical engineers, electrical engineers, consultants, state officers, bankers etc. have to work together. In the professional life, engineer generally has to overcome very sophisticated problem during execution stage of the project. In order to solve those problems, it is necessary to know point of view of different professions. Those problems can be solved by engineers have broader perspective. This perspective can be given during education period. However, existing education system leads students to go into technical details of specific problems not common in practice. In this paper, it is aimed to investigate the lack of interdisciplinary studies during university education in Turkey and the world. In addition to this in order to decrease unemployment rate in university graduate students, the qualification of the students must be increased. This study mainly focused on the two aspects: theoretical discussions and exemplification of practical concern of interdisciplinary education.

Keywords: Education, interdiscipliner, engineering, architecture, university

1 INTRODUCTION

Engineering is a field of broad scope that ranges from the design and construction of structures, roadways and pollution control processes to the management of natural and engineered resources. The breadth of the field has led to division into a number of subdisciplines including structural engineering, geotechnical engineering, transportation engineering, environmental engineering, materials, construction management, and water resources engineering. As the world's dependence on technology increases, there is no question that the demand for engineers is increasing. Engineering is a key component of innovation and technological society. That's why students generally prefer engineering faculties for BSc. education in both developed and developing countries as shown in the Figure 1.

As shown in the Figure 1, the increasing industrial growth rate is parallel to engineering degrees in China. In order to sustain industrial growth rate, it is necessary to educate more students as engineer.

According to OECD education data for 2007, engineering popularity in developed countries is higher than OECD countries as shown in Figure 2. However natural science is generally more popular in OECD countries in compared with developed countries.

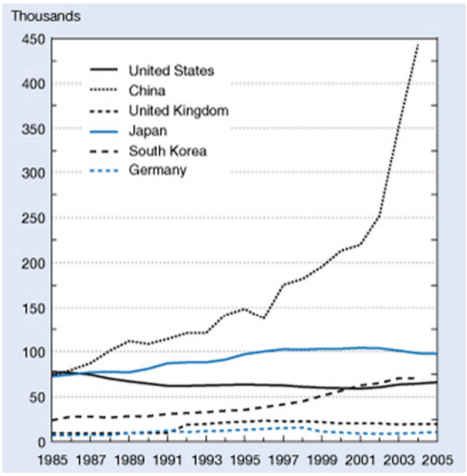


Figure 1. The university engineering degrees by selected countries, 1985-2005.

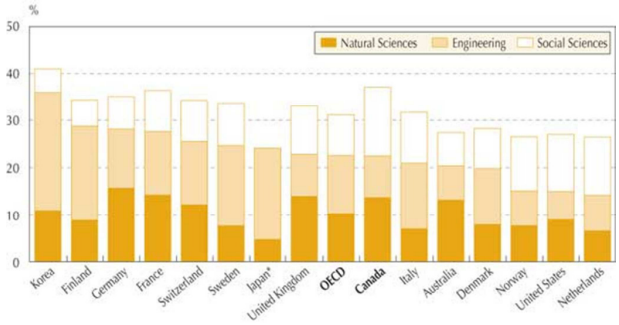


Figure 2. Science and engineering degrees as a percentage of new degrees for selected OECD countries [1]

Especially for developing countries, civil engineering is one of the most important engineering branches. Because developing country has to construct many buildings, schools, factories, roads, tunnels, dams, power plants etc. Civil engineering draws on a number of science disciplines including chemistry, physics, ecology, geology, microbiology, material science, economics and mathematics, and statistics. Civil engineering is about community service, development, and improvement, planning, design, construction, and operation of facilities essential to modern life, ranging from transit systems to offshore structures to space satellites. Civil engineers are also problem solvers, meeting the challenges of pollution, traffic congestion, drinking water and energy needs, urban redevelopment, and community planning.

A collaborative project between the faculties of computing and education, beginning with a study to identify the problems experienced by students and staff, has resulted in a series of modifications to the structure and delivery of the subjects and the development of a tutor training course almost impossible to take into account reliably only a century ago the nature of civil engineering education was entirely different than today. It was more like art than science, focused on creativity and leadership and on building a holistic, or qualitative, understanding of our profession. The available knowledge was mostly heuristic in the form of decision rules acquired through centuries of practice following the master-apprentice paradigm. To-day, the focus is mostly on the analysis, on building quantitative understanding and numerical optimality, as it is in science. The civil engineering knowledge is only partially heuristic, over the last century it has been supplemented by all kinds of mathematics- and physics-based theories, including complex mathematical models. Civil engineering became a science, but at the same time becoming painfully aware that the price for this progress is the loss of our creativity and the excessive focus on the quantitative aspects of this profession. This shift from art to science has ultimately caused that civil engineers are inadequately prepared to deal with complex challenges of the 21st Century, which require novel solutions produced by out-of-the-box bold thinking [2].

Civil engineers are rarely perceived as proactive and creative leaders. They are usually seen as reactive technologists and followers. Subsequently, public attention moved to other areas of technology, mostly to Information Technology. As a result of that, a number of undesired phenomena occur today, including growing stagnation in civil engineering, deteriorating infrastructure, reduced infrastructure spending, etc for developed countries. This situation has also caused that the best and brightest students choose not to study civil engineering and a lot of enthusiasm, creativity, and prides are lost in the process. In addition to this, students are doubtful about their level of practice and capacity of problem solving for professional life. Because in professional life, they have to deal with many different professions and have to solve very sophisticated problems. They also have to manage and coordinate a site which is full of human. Recent years, the unemployment rate in university graduated young people has been increasing in developed or developing countries (i.e. %42 in Spain). In order to decrease unemployment rate in university graduate students, the qualification of the students must be increased. The interdiscipliner study practice is one of the important experience and reason for employers. That's why they have to have basic terminology about management, human relations, psychology, finance, design, mechanics, electricity etc.

2. EXISTING EDUCATION SYSTEM

Over the years there have been recommendations from employers and various technical and professional organizations to revise the engineering curriculum to ensure that students are prepared for the professional practice of engineering [3]. Practicing engineers and educators have also in-

dicated that they are not completely satisfied with the average engineering program in the United States [4], [5], [6]. One aspect that appears to be missing is the development of practical experience and leadership skills in students. However, many civil engineering departments have encouraged collaboration with industry to increase the practical outlook of their students [7]. It has also been written that professors should be rewarded for being involved in industry. In this manner, it should be possible to incorporate technology into the curriculum at all levels [8]. Today, there is a movement in engineering education toward exposing students at all levels to more realistic and practical design experiences. For example, at some schools client sponsored design-project courses are required for both undergraduate and graduate students [9]. Here, the involvement of practitioners is vital since a small number of traditional faculties, according to some authorities, may lack the depth of practical project experience and judgment that engineers in business and industry tend to possess [10]. Financial need often forces individuals to be part-time students. If the work is in the engineering area, the arrangement allows for better skills integration on the job [11]. Another strong benefit is to bring practical concepts and ideas into the class-room, which can be shared with others [12]. In this regard, it has been found that students involved with internship programs have been made aware of various career options, increased their engineering and technical knowledge, improved their communication skills and have obtained contacts for future employment [13]. In addition, it has been observed that some organizations appear to prefer hiring new employees who have had an intern, co-op, or other work experience with the firm. This tends to reduce the possibility of a mismatch between the new engineer and the organization and smooth's the road between school and work [14].

As shown in the Figure 3, the most preferred engineering types are mechanical, electrical and civil engineers. Many engineers and other disciplines have to work together in large scale projects. For example, civil engineers, architects, electrical engineers, mechanical engineers, archeologist, paleontologist, marine biologist etc. has worked and cooperated during the Marmaray project (tunnel project under Bosporus). This interdisciplinary working culture must be given in university life.

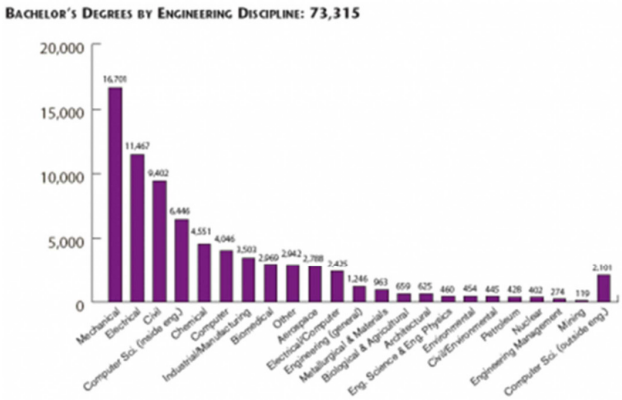


Figure 3. The most preferred engineering types for BSc.

During under graduation education, those engineering students have to learn to work together. They have to understand their technical language and way of thinking. Otherwise they will learn to understand each other during their professional life by losing money and time [15].

Turkey has been envy of the whole world with its tremendous success in coming out of the recent economic crisis of 2008 relatively unscathed. What is even more commendable is that the recent growth in GDP is accompanied by a record decline in inflation. Nevertheless, unemployment has long been an intractable quandary in the country. As shown in Figure 4, the initial effect of the current economic crisis is a rapid jump in the unemployment rate, which has taken a reverse course, especially since mid-2009.

However, what is worse for the unemployment problem of the country is the inconsistency between unemployment and the educational attainment for the young people who are the most concerned about the said relationship. Table 1 reveals that among the holders of a degree, the highest portion of the unemployed is college graduates. As a matter of fact, there is a positive relationship between the level of the degree held and unemployment rate; the higher the degree, the higher the unemployment.

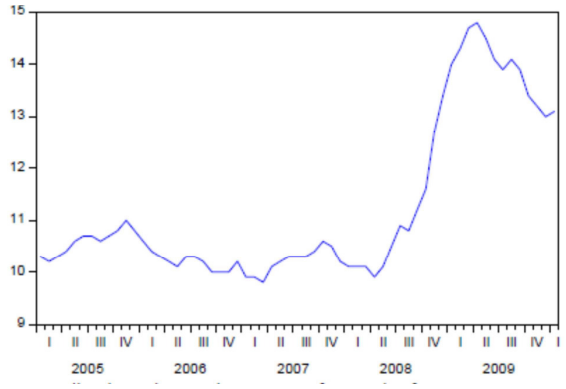


Figure 4. Seasonally adjusted unemployment rate for people of 15+ age. Source: TUIK.

A curious observation in Table 1 is that vocational school graduates are much more successful than their counterparts among ordinary high schools. Even more curiously, vocational school graduates have achieved lower unemployment rates than college graduates. To put the issue in perspective, we observe in Table 2 that the unemployment rate declines in the US as level of the degree held. Additionally, Turkey seems to be lagging behind in many aspects compared to the European Union countries [16].

In sum, there is a peculiar employment situation in Turkey with respect to the educational attainment. Bleak prospect for higher degree holders may discourage young people from pursuing more and more education, but instead they may choose to seek opportunities in the job market with less education. In Europe, however, young people may refuse to go to school as they can find jobs with less employment

Table 1. Turkish Youth Unemployment Rate According to Education Levels

year	primary	Middle	high	vocation	college
2000	8.3	13.7	20.6	20.8	28.3
2001	10.6	17.7	24.0	25.5	30.7
2002	12.4	20.7	26.8	28.0	38.3
2004	13.4	19.6	26.6	29.3	39.8
2005	14.1	19.2	25.3	25.6	30.5
2006	14.7	17.9	25.2	21.1	27.2
2007	14.7	19.3	23.5	22.6	28.5
2008	14.3	18.7	25.0	20.8	29.8
2009	17.9	21.6	30.6	27.6	33.2
2010	14.9	10.9	27.2	23.1	32.5

3. METHOD

The proposed methods are given below. In order to establish interdisciplinary culture during education life, this period and methods can be divided into stages.

3.1 The Interdisciplinary studies during bachelor stage

The Bachelor of Science (BSc.) period is the most important stage to teach student how to perform interdisciplinary studies. In order to establish a common language, students have to take courses together. Therefore the university system must be flexible and let students to take courses from other departments. The most important point in this concept is the content of the courses should be reviewed and updated for students from different departments. The development of the content and of the teaching material for courses is vital. Teaching materials have been prepared in digital form following the e-learning standards by experts from various engineering department. These courses can be divided into two categories: elective courses and mandatory courses. Both category courses must have project part which should be prepared by student groups. These student groups should be established by students from different departments. During the preparation phase of the project, students will learn technical terminology of other engineering branches. Students should experience how interesting and challenging engineering disciplines can be, how to apply theoretical principles and equations to practical experience and, most importantly, how engineering can make them think.

The second tool to teach interdisciplinary culture is the encouragement of double major programs. The schedule, content and total credits of the programs should be arranged to make those double major programs attractive as much as possible. When student graduate from a double major program of engineering departments, he or she will have broader perspective compared with one department graduation.

The third tool is to encourage students to join national or international project contests. These

types of contest exist in every country. Many large scale companies and officials foundations financially support this kind of contests. These organizations give an opportunity to the students to study together for a specific technical target. These interdisciplinary studies teach students how to work together.

The fourth and last tool is to develop new elective courses which will be given by construction professionals. The aim is these courses to share the experience of construction professionals with students. Students need to hear about professional life and measure their knowledge level is enough or not.

3.2 The Interdisciplinary studies during graduate stage

The graduate stage can be divided into two phases. The first phase is course phase. The second phase is thesis phase. During the first phase, students should take courses from different departments. In order to success this concept, the content of the courses should be reviewed and updated for students from different departments. For the second phase, the subject of the thesis should be interdisciplinary. This type of thesis can be supported by National Science Foundations. Instructors have a key role for graduate students. Therefore university professors should believe the importance of interdisciplinary and they should have practice background and relations with industry. These academics should have publication in this area. In addition to this, for PhD, a quota should be given for students from different departments.

As alluded to in Figure 5, the theory of business and engineering to yield entrepreneurs, who are not only compelled to look for jobs in the public sector, but potentially capable of generating employment for others. As shown in the Figure 5, our aim is not to dwell deeply on the depth of each field, i.e. business and engineering, but instead to explore more on the opportunities to expand the integration along the different fields (width axis). A constant interaction between theory and practice during the program to refine skills in business management and related issues [16].

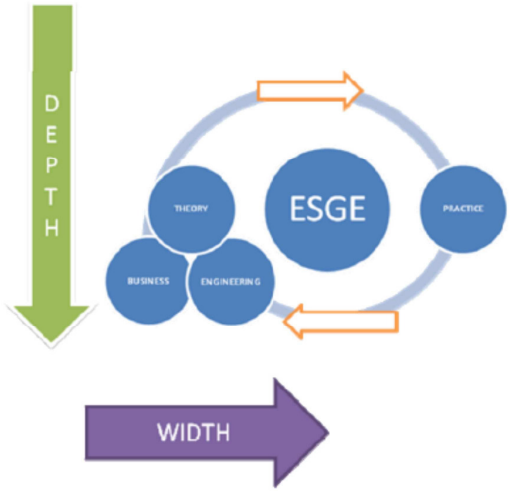


Figure 5. Relation between theory and practice

The educational institutions, in order to avoid being aloof to changes in society, have to turn out graduates with the necessary skills to deal with the issues of the modern world. Graduates of our program are expected to be managers for the modern world.

4. RESULTS AND DISCUSSIONS

As explained previous sections, engineers are working for large scale project throughout world. These projects can be constructed by different professions. That's why different professions should communicate with each other which lead save time and budget of projects. The proposed new engineering education paradigm is feasible, but its implementation is difficult and expensive. The globalization of civil engineering education is becoming a fact. A number of international research programs focused on engineering education have already been initiated. Most importantly, the issue of the today's focus on the interdisciplinary aspects of engineering must be addressed. It should be based on a balance of technical and social sides, or, more specifically, on a balance of teaching both qualitative and quantitative knowledge and skills with prominent focus on teaching inventive design in the context of engineering problem. That could be done through the integration of different departments and problem solving with a number of courses offered through the entire program. Global challenges require global action and cooperation. Therefore, the nature and extend of challenges simply needs global cooperation, par-

ticularly that the resources of a single country, not to mention a single university, are grossly inadequate for such an important, difficult, and huge task. How such cooperation could be organized is still an open question. However, the products most likely should be in the form of teaching modules, ultimately developed as multimedia intelligent tutoring systems. In addition, the interdisciplinary activities should enhance the skills required by engineering students for a successful career involving the design and management of engineering and construction projects.

5. REFERENCES

[1] OECD Education database, 2007, www.oecd.org/ff/?404;http://www.oecd.org.../education/database

[2] Arciszewski, T., Civil Engineering Crisis, **ASCE Journal of Leadership and Management in Engineering**, pp. 26-30, Vol. 6, No. 1, January, 2006.

[3] Engineering education for a changing world, **American Society for Engineering Education (ASEE)**,1994, Washington, D.C.

[4] 'New course curriculum will emphasize design.', 1998, **Civil and Environmental Engineering at MIT**, 12(3), 1–2.

[5] Mendelsohn, R., 1998, 'Teamwork-The key to productivity, **J. Manage. Eng.,** 14(1), 22–25.

[6] Sabatini, D. A., 1997, 'Teaching and research synergism: The undergraduate research experience, **J. Prof. Issues Eng. Educ. Pract.,** 123(3), 98–102.

[7] Tener, R. K., 1996, 'Industry-University partnerships for construction engineering education**, J. Prof. Issues Eng. Educ. Pract.**, 122(4),

[8] Farr, J. V., 1997, Engineering education: The challenge of reform, **J. Manage. Eng.,** 13(6), 3–4.

[9] Jones, S. A., and Houghtalen, R., 2000, Using senior design capstone as model for graduate education, **J. Prof. Issues Eng. Educ. Pract.,** 126(2), 83–88.

[10] Kerkes, D. J., 1995, Arguments for recruiting faculty from business and industry, **J. Prof. Issues Eng. Educ. Pract.**, 12(4), 242–245.

[11] Hensey, M., 1998, **Graduate education in the workplace**, J. Manage. Eng., 14(2), 6–7.

[12] Bernold, L. E., 2000, Typical lectures fail students, **Eng. News-Rec**., 23(67), 244.

[13] Croissant, J. L., Ogden, K., and Ogden, G., 2000, 'Teamed internships in environmental engineering and technology: Project report, **J. Eng. Educ**., 89(2), 111–114.

[14] Goldberg, D. E., 1998, The human side of engineering, **The Bulletin of Tau Beta Pi**, Knoxville, Tenn., 71(3), 1–3.

[15] Güney, D., İnşaat Mühendisliği Eğitiminde Disiplinlerarası Çalışma Eksikliğinin Giderilmesi, **1. İnşaat Mühendisliği Eğitimi Sempozyumu**, 6-7 Kasım 2009, Antalya, Türkiye.

[16] Genc, I. H., 2009, Designing an Integrated Business Curriculum with Students' Success in Mind: An Evaluation within the Context of the IBC Program at the University of Idaho, **The International Journal of Management Education**, 7(3): 81-86.

Tablet PC Support of Students' Learning Styles

Shreya KOTHANETH
Industrial and Systems Engineering, Virginia Tech
Blacksburg, VA 24060, USA

Ashley ROBINSON
Computer Science, Virginia Tech
Blacksburg, VA 24060, USA

and

Dr. Catherine AMELINK,
Division of Student Affairs, Virginia Tech
Blacksburg, VA 24060, USA

ABSTRACT

In the context of rapid technology development, it comes as no surprise that technology continues to impact the educational domain, challenging traditional teaching and learning styles. This study focuses on how students with different learning styles use instructional technology, and in particular, the tablet PC to enhance their learning experience. The VARK model was chosen as our theoretical framework as we analyzed responses quantitatively and qualitatively of an online survey. Results indicate that the tablet PC does cater to all learning styles.

Keywords: Learning Styles, Tablet PC, VARK, Instructional Technology, and Higher Education.

1. INTRODUCTION

In the context of rapid technology development, it comes as no surprise that technology continues to impact the educational domain, challenging traditional teaching and learning styles. If employed correctly, instructional technologies, in particular, have been found to benefit the learning experience by aiding conceptualization, applying knowledge, and facilitating dialogue [1]. Research has shown that it is important to understand learning styles and profiles of students in order to have the most effective teaching and learning experience. Studies have found that multimedia technology have a positive effect on a classroom of students with different learning styles by neutralizing differences in performances [2]. This study focuses on how students with different learning styles use instructional technology, and in particular, the tablet PC to enhance their learning experience.

2. BACKGROUND

Learning Styles

The concept of learning styles originated in the 1970s. It has been defined as different approaches to learning. Some researchers argue the most effective instructional technique is to first identify learning styles and then adapt their instructional method to each learning style [3]. Others suggest that there is no single efficient teaching method; educators should instead use a variety of instructional techniques [4,5]. In this technological era, it is important to understand the relationship between learning styles and the use and acceptance of technology.

Various instruments have been developed to understand learning preferences. The Learning Style Inventory [6], was developed by Kolb, to identify and categorize participants into four learning styles; Diverging, Assimilating, Converging, and Accommodating [6]. The Dunn and Dunn Learning Style Inventory [3] is based on the notion that students have learning style preferences which are divided into twenty-one variables that affect learning in five categories: Environmental, Emotional, Sociological, Physiological, and Psychological [3]. The Myers-Briggs Type Indicator [7] divides learning preferences into four dichotomies, which results in sixteen possible psychological learning types, including Extroversion versus Introversion, Sensing versus Intuition, Thinking versus Feeling, and Judgment versus Perception [7].

One of the most popular models is Fleming's VARK model [5], which divides learners into four categories: Visual (V), Auditory (A), Reading/writing (R), and Kinesthetic (K) (See Figure 1). Visual learners are those who learn best with visual artifacts. Auditory learners are those that learn with oral stimulations. They learn by

talking and listening. Reading or Writing learners prefer printed words to gain knowledge. Kinesthetic learners are those who learn by experience. They learn by real world examples and by application. Learners can also be multi-styled, making use of two or more learning styles [5]. Since this model is extremely popular and widely used, we have chosen this model as our theoretical framework.

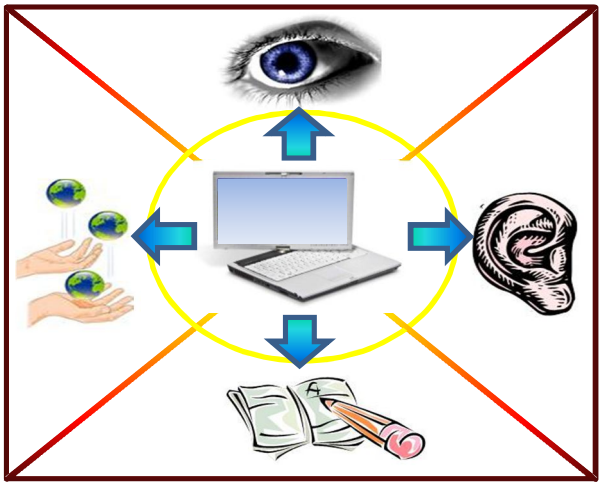


Figure 1: Conceptual model depicting relationship between students' learning styles (VARK Model) and the use of tablet PC

Instructional Technology

The rapid ascent in computer technology and the Internet has contributed to increase in using different media for education. Instructional technologies can range from using software like PowerPoint, to hardware like laptops and tablet PCs in the classroom. These media can improve the teaching and learning experience. For instance, Moore [8] found that the use of instructional technologies led to an increase in three types of interactions: between the students and learning material, students and instructor, and among the students. Various technologies have also been identified as useful for students who belong to Generation Y, who are characterized by their high use and dependency on technologies [9]. Some researchers go so far as to claim that teaching will eventually only be conducted by technology and multimedia instructional techniques as they are far superior to standard instructional techniques [10].

Tablet PCs

The tablet PC, in particular, is becoming a very common tool in higher education. For instance, Pennsylvania State University [11], University of Toronto [12], and University of Alaska Anchorage [13] are all examples of universities who have starting using the tablet PC in some of their courses. The tablet PC is like a standard laptop, with a keyboard for typing as well as a stylus for scribing. The screen can be rotated and collapsed, so that a stylus can be used to make handwritten notes and drawings in a similar fashion to pen and paper. The addition of the stylus is the main difference between the tablet PC and other slate-like devices such as the iPad. Tablet PCs afford rich graphics that aid in visualization. The built in speaker and microphone not only enable richer audio presentations, but affords audio recording for future playback. The electronic inking (e-inking) capability of the tablet, which enables users to write on the tablet with a stylus as if it were an ink pen, provides open-ended note-taking capabilities. The slate conversion features enables users to read/write on a flat surface, simulating notebook reading/writing. There is also tablet PC based software that enables students to create and interact with their learning environment [14].

With the increasing trend of tablet PCs and slate like technologies, there is a need to understand the relationship between learning styles of students and use of the technology in the classroom. The College of Engineering, at Virginia Tech, started the Tablet PC Initiative in 2006, which requires all incoming engineering freshman to purchase and use a tablet PC in class. All classrooms have access to the University's wireless network. This study uses the College of Engineering at Virginia Tech as a forum to understand better how the tablet PC is used by different types of undergraduate engineering students. Using software that enhances Tablet PC features such as Microsoft OneNote [15] and DyKnow [16] are common among Virginia Tech students. Microsoft OneNote is software that supports extensive note taking, note sharing, audio recording, e-inking, and collaboration capabilities. DyKnow is classroom interactive classroom software that contains such features as polling, lecture slide annotations, audio recording, e-inking, and collaboration.

3. METHOD

An online survey was sent to all undergraduate students enrolled in the College of Engineering at Virginia Tech, after approval by the Institutional Review Board (IRB) as part of the engineering curriculum assessment process. This questionnaire, designed using expert review, elicited information on how they study, organize material, use the tablet PC among other things, and took approximately 15 minutes to complete. To gather qualitative data, we also asked students the following question: "Consider how the Tablet PC has been used by your different instructors. In your experience, what activities are best supported with the Tablet PC?" The survey obtained a usable response rate of 18% with a total of 1090 students. All the students were enrolled either in or after Fall 2006, which implied that they were affected by the tablet PC initiative.

Flemings' VARK model served as the theoretical framework for the analysis. Of the all the questions sent in the original survey, seventeen questions were selected on their ability to elicit use of the tablet PC relating to the four learning styles. Some items used a four-point Likert scale with anchors 1-Never, 2-Rarely, 3-Occasionaly, and 4-Frequently. Other items used a five-point Likert scale with anchors 1-Never, 2-Rarely (once or twice a semester), 3-Sometimes (several times a semester), 4-Frequently (five or more times a semester), and 5-Very frequently (almost every week). The questions were coded as V, A, R, or K. See Table 1 for more details.

Visual questions were based on students' use of charts and diagrams to better understand course material. Auditory questions related to making use of the tablet's audio recording abilities to record lectures and also discussing the course material with their peers. The reading and writing questions captured the student's use of the tablet's e-inking feature to take and review notes. Kinesthetic questions were based on students' interactivity with the course material.

Data from the survey was analyzed using descriptive statistics. In addition, qualitative responses of the students were analyzed using content analysis, specifically thematic coding, to gain a further understanding of how students with different learning styles used the tablet PC.

4. RESULTS

Quantitative Analysis

Our results, in Table 1, indicate that tablet PC usage in the classroom addresses all learning styles. Highest mean scores were reported for using the e-inking capabilities of the tablet PC to help illustrate points made in class and mark slides, and using web-based sources to apply concepts in class.

Qualitative Analysis

Visual Learning: Several students indicated that they liked using the tablet PC to create visual representations of their notes. They reported that it helped with creating diagrams and for sketching assignments. Students also reported that the tablet PCs allowed professors to enhance their teaching experience by adding visual artifacts to their notes. They said, "Used e-ink to write out examples or show drawings and such. Drawings are best supported."

Auditory Learning: Software like DyKnow and OneNote was used to record lectures which were helpful for review. According to one student, "Group brainstorming or planning sessions work well with a shared OneNote page." Another student commented, "DyKnow and OneNote are both just lovely. I am a huge DyKnow fan. I like how the pen strokes can be replayed using DyKnow and the voice recording is a very useful tool as well."

Learning Style	Mean	SD
Visual		
* E-ink to create diagrams	2.16	1.09
* Shared electronic whitespace with other students	1.56	0.83
Helped illustrate points made in class	3.31	0.95
** I made simple charts, diagrams, or tables using the Tablet PC to organize course materials	2.22	1.23
Auditory		
* Shared notes/slides with other students	1.94	0.98
* Shared electronic whitespace with other students	1.56	0.83
* Audio recording of lectures/discussions using OneNote	1.23	0.59
Reading/Writing		
* E-ink to mark slides provided by the instructor	2.39	1.17
* E-ink to take notes using OneNote	2.30	1.08
* Imported web-based information into notes	2.00	1.10
* E-ink to take notes with another program	1.93	1.06
* Special note take capabilities of OneNote	1.69	0.97
** I study by reading my notes over and over again	2.90	1.24
** I make lists of important items for this course and memorize the lists	2.26	1.18
Kinesthetic		
* To respond to interactive class exercises using polling/voting	1.77	0.88
* To respond to interactive in-class exercise using written responses	1.74	0.91
** I try to apply ideas from web-based sources to other class activities such as lecture and discussion	2.42	1.14

Table 1: Questionnaire Result Means and Standard Deviations. **Used a five-point Likert scale with anchors 1-Never, 2-Rarely (once or twice a semester), 3-Sometimes (several times a semester), 4-Frequently (five or more times a semester), and 5-Very frequently (almost every week). *Used a four-point Likert scale with anchors 1-Never, 2-Rarely, 3-Occasionaly, and 4-Frequently.

Reading/Writing Learning: A majority of students indicated that the tablet PC was useful for note taking. Tablet PCs were also found to be useful in classes that involved a lot of equation writing. One student reported, "Tablet PC is good for taking notes. Taking notes on the actual slides that a professor is teaching from is much more beneficial than simply taking notes in paper

notebook. The tablet is also convenient in that it allows you to take notes without having to print out the lecture slides."

Kinesthetic Learning: Students indicated that the ability to swivel the monitor helped them to collaborate with each other. Students also liked that they could transfer their drawings and sketches easily using the tablet PC. The use of interactive software, such as DyKnow [16] and OneNote , in class made the class a lot more engaging and students reported enjoying taking polls in class. Students said that, "Activities involving polls and giving responses to the teacher are best supported with the Tablet PC."

5. DISCUSSION

The quantitative and qualitative analyses reveal that the tablet PC can be beneficial to all learning styles. Visual and reading/writing learning style features were used the most. This can be attributed to the fact that lectures are still the most popular teaching styles thus automatically cater to auditory learners. The main benefit of using a tablet PC in class is to supplement the lectures with visual and reading/writing artifacts. Reading and writing are natural affordances of the tablet due to the slate conversion option and the stylus.

The qualitative analyses revealed that students use additional features such as recording the lectures. These features will also be useful in an online course, where the instructor is remote. Software like DyKnow [16] and OneNote [15], can be used to make the class more interactive and thus be helpful for kinesthetic learners.

It is important to acknowledge that the lower mean scores (means below 3.00 for a 5-point Likert scale and below 2.00 for a 4-point Likert scale) can be attributed to what's also been biggest barrier to the Tablet PC Initiative at Virginia Tech, the lack of continued use of the tablet PC by faculty members in all the engineering classes. While tablet PCs have been extensively used in the freshman classes, their use in class by faculty members drop extensively in the senior classes.

One limitation of this study is that the survey was designed to gain an understanding on the use of the tablet PC but not specifically on the learning styles of the students. In addition, the hardware of the tablet PC could also aid various learning styles; however, we did not investigate this question. We recommend that questions related to the tablet PC's hardware (e.g. touchpad/mouse, keyboard, size, etc.) be included in any future study.

6. CONCLUSION

Our results show that the tablet PC is useful for students who possess any of the VARK learning styles. However, success of tablet use in the classroom ultimately depends on efforts to prepare engaging lectures using the technology [17]. This is apparent in the lower kinesthetic means, which heavily depends on interaction of students in the lecture. If instructors do not implement interactive features such as polling and short response questions, kinesthetic learners will be at a disadvantage, because they 'learn by doing' and will not have the opportunity to use the technology to engage with the course material. The lecture should also be visually engaging and also be made available to students so that students can read and annotate their notes. Instructors can audio record their lectures using features of the tablet PC and let students access it for future playback. While it is evident that the tablet PC can cater to all learning styles, we envision that they will better support VARK learning styles as instructors as well as students become more comfortable with tablet PCs and their features.

7. REFERENCES

[1] J. McKendree, K. Stenning, T. Mayes, J. Lee, and R. Cox, "Why observing a dialogue may benefit learning," *Journal of Computer Assisted Learning*, vol. 14, 1998, pp. 110-119.

[2] F. Karakaya, T.L. Ainscough, and J. Chopoorian, "The Effects of Class Size and Learning Style on Student Performance in a Multimedia-Based Marketing Course," *Journal of Marketing Education*, vol. 23, Aug. 2001, pp. 84-90.

[3] R.S. Dunn and K.J. Dunn, *Teaching students through their individual learning styles: A practical approach*, Reston, Va: Reston Publishing Company, 1978.

[4] M. Sprenger, *Differentiation through learning styles and memory*, Thousand Oaks, CA: Corwin Press, 2003.

[5] N.D. Fleming, "I'm different; not dumb. Modes of presentation (VARK) in the tertiary classroom," *Research and Development in Higher Education, Proceedings of the 1995 Annual Conference of the Higher Education and Research Development Society of Australasia (HERDSA), HERDSA*, 1995, p. 308–313.

[6] D.A. Kolb, *Individual learning styles and the learning process*, Cambidge, Ma: MIT Press, 1971.

[7] I.B. Myers and M. McCaulley, *Manual, a guide to the development and use of the Myers-Briggs type indicator*, Palo Alto: Consulting Psychologists Press, 1985.

[8] M.G. Moore, "Three types of interaction," *The American Journal of Distance Education*, vol. 3, 1989, pp. 1-6.

[9] W. Strauss and N. Howe, *Generations: the history of America's future, 1584-2069*, New York: Morrow, 1991.

[10] J. Gandz, "The death of teaching: the rebirth of education," *Ivey Business Quarterly*, vol. 62, 1997.

[11] S.G. Bilen, D. Lee, J.I. Messner, H.T. Nguyen, T.W. Simpson, A.A. Techastassanasoontorn, and R.F. Devon, "Tablet PC Use and Impact on Learning in Technology and Engineering Classrooms: A Preliminary Study," *Workshop on the Impact of Pen-based Technology*, West Lafayette, IN: 2008.

[12] M. Stickel, "Effective Use of Tablet PCs for Engineering Mathematics Education," *ASEE/IEEE Frontiers in Education*, Saratoga Springs, NY: 2008.

[13] K. Mock, "Teaching with Tablet PCS," *Journal of Computing Sciences in Colleges*, vol. 20, 2004, pp. 17-27.

[14] P. Hamilton and E. O'Duffy, "Digital education usage models for the classroom of the future," *Proceedings of the 4th International Conference on Virtual Learning*, 2009.

[15] Microsoft, "OneNote 2010," 2011.

[16] Dynamic Knowledge Transfer, "DyKnow," 2010.

[17] M. Stickel and S.V. Hum, "Lessons learned from the first-time use of tablet PCs in the classroom," *Frontiers in Education Conference, 2008. FIE 2008. 38th Annual*, IEEE, 2008, p. S1A.

What Do Fourth Year Engineering Students Know About Sustainable Development?

Iacovos NICOLAOU

School of Civil Engineering and Building Services,

Dublin Institute of Technology, DIT

and

Eddie CONLON,

Department of Engineering Science and General Studies, DIT,

Dublin, Dublin 1, Ireland

ABSTRACT

This paper presents data from a project which aims to determine the level of knowledge and understanding of engineering students about Sustainable Development (SD). The data derives from a survey completed by fourth year engineering students in three Irish Higher Education Institutions. This paper is part of a larger study which examines the relationship between students' and teachers' understanding of SD. The results from the survey show that fourth year engineering students have a discipline-led conception of SD. The majority of the participants fail to acknowledge the complexity of the concept and focus only on environmental protection. Their knowledge of legislation relevant to SD and the social aspect of sustainability is deficient.

Keywords: Engineering Education, Sustainable Development

INTRODUCTION

This paper will focus on the knowledge and understanding of fourth year engineering students about Sustainable Development (SD). The data comes from a larger study, which is still in progress, which is examining the relationship between the understanding and knowledge of engineering students and their teachers about SD.

There is now a requirement that engineers practice and promote the principles of SD. The mission of Engineers Ireland (EI), the professional body for engineers in Ireland now includes a commitment to promote SD. It states: "Our members serve society through the highest standards of professional engineering. We seek to improve the quality of life for all, creating prosperity and adding value through innovation and the promotion of health, **and sustainable development"** (emphasis added)**.**

It is widely accepted that engineers can play a key role in delivering sustainability [1] [2] with some arguing that they are uniquely placed to take a lead in moving towards sustainability [3]. As Johnston puts it "engineers really are necessary to make sustainability work" [4]. Moreover, Ashford argues that a specific focus should be given to engineers to achieve SD since they drive any kind of development [2].

However, existing research shows that engineering students do not understand the complexity of the concept seeing it purely as linked to environmental issues without understanding the social dimension of SD [1] [5] [6] [7] [8].

The motivation for this project was drawn from the lack of research regarding engineering education for SD in the Irish context. This is despite the requirement, set out in the Code of Professional Ethics of Engineers Ireland, for engineers to practice and promote the principles of sustainable development and the accelerating demands for engineers' competency on sustainability issues.

The work reported in this paper seeks to establish the extent to which Irish final year engineering students share the narrow view of the concept as reported in the literature. This paper seeks to build on two pieces of research that have investigated what engineering students know about SD.

SD origins and definition

Sustainable Development (SD) is a concept that was first introduced in the 1980's. The concept was presented in the World's Conservation Strategy (1980). The three main pillars that constitute SD are the environmental, the economic and the social aspect. SD is most commonly defined by the World Commission on Environment and Development, the Brundtland report [9] as *"development that meets the needs of the current generation without compromising the ability of future generations to meet their own needs"*.

What do engineering students know about SD?

As stated above this study draws on two pieces of research which have investigated engineering students' knowledge of SD.

In the first, Carew and Mitchell [6] conducted a qualitative investigation of students' understanding of the concept in the University of Sydney. The data showed that students' understanding of SD was very broad with no evidence that they understood the complexity of the concept. They classified students' descriptions of SD. Their classification was based on an analytical framework for mapping variation in student

conceptions called the Structure of Observed Learning Outcomes (SOLO) taxonomy introduced by Biggs and Collis [10]. The classes used for classification are in Table 1.

SOLO Classes	Features of Sustainability typical for each stage
1. Pre-structural	Either did not know what sustainability was or provided a non-specific response
2. Uni-Structural	Provided one definitive example of something concrete and relevant to SD
3. Multi-structural	Provided two or more different examples of things relevant to SD
4. Relational	Constructed a cohesive statement about SD by relating two or more things relevant to SD
5. Extended Abstracts	Constructed a cohesive statement about SD by relating two or more things relevant to SD and provided evidence of critical/creative thinking or ethical judgment

Table 1: Carew's and Mitchell SOLO taxonomy [6]

The results of the classification showed that 65% of the responses were classified as pre-structural and uni-structural which corresponds to answers that either did not know what sustainability is; or had a very vague perception of the concept. 18% of the responses were classified as multi-structural which corresponded to answers where an attempt to define sustainability was based on the combination of two relevant aspects of the concept such as environmental protection and future needs. The remaining 17% was allocated to relational and extended abstract classes which included responses showing knowledge evidence of critical thinking and ethical responsibility. Carew and Mitchell argued that the results are concerning since the respondents were at the third year of their degree which assumes that some modules relevant to SD were completed by them. They said that curricula should be improved in order to facilitate students that have a pre-structural conception and assist them to develop their knowledge, while the same system would provide opportunities to more knowledgeable students to further explore the concept.

In the second investigation, Azapagic and her co-researchers [1] carried out an international quantitative survey of engineering students in order to determine their level of knowledge and understanding of SD and identify knowledge gaps. The survey was distributed to 21 countries and a total of 3134 students completed the questionnaire. The data showed that engineering students tend to connect SD with environmental issues and neglect the other two pillars of the concept (economic and social). The results show that students' knowledge and understanding of SD was particularly low. Substantial knowledge gaps were found in regard to SD legislation, policy and standards, SD social issues and several environmental issues such as loss of biodiversity and salinity.

Azapagic also found that respondents thought that SD is an important concept for them; and more important for future generations. Their results did not show any significant difference when variables such as gender, discipline and year of study were examined. Nevertheless, participants from Sweden, Vietnam and Germany were distinguished from the sample with a higher level of knowledge of SD compared with other participating countries such as UK, Central Europe and the U.S.A. Based on the above, Azapagic [1] argues that engineering students' level of knowledge and understanding of SD is "not satisfactory" and that deficiencies in engineering education should be minimized in order to adequately educate engineering students and close the knowledge gaps mentioned above.

Carew's and Mitchell [6] and Azapagic's et al [1] work showed major gaps in engineering students' knowledge and understanding of SD.

This study draws from those two projects as a framework to determine the level of knowledge and understanding of fourth year engineering students of SD in the Irish context. The results will be presented in the results section where comparisons will be made between our results and those from Azapagic et al [1] and Carew and Mitchell [6].

METHODOLOGY

The data reported in this paper comes from a survey of fourth year engineering students across a range of engineering disciplines in a number of Irish higher education institutions (Table 1). One of the institutions is ranked as one of the top 56 higher education institutions that promote SD in engineering education [11]. The institutions are located in three different cities. One of them is a traditional and long standing university; one is a new university having achieved university status in 1989; while the third is an institute of technology offering programs from traditional apprenticeships to doctorates.

Degree / *Inst.*	Civil Eng	Mechanical Eng	Chemical Eng	Structural Eng	Building Services
Inst. 1	× (25)	× (29)		× (26)	× (25)
Inst. 2	× (23)		× (7)		
Inst. 3		× (8)			

Table 2: The three Institutes and the degrees surveyed. Numbers in parentheses illustrate response from each degree and institute.

A questionnaire was designed drawing on that used by Azapagic et al. [1]. Carew's and Mitchell [6] open-ended approach was incorporated asking students to state, in their own words, their understanding of SD. However, the two questionnaires were not identical. Additions were made in order for the questionnaire to be appropriate for the Irish context. SD principles were drawn from Ireland's SD Council, Comhar [12] while SD tools were identified in Mulder's book on SD for engineers [13].

The main research question "**What is the level of knowledge and understanding of fourth year engineering students of SD**" was utilized to generate subsequent questions that were used in the survey. The questionnaire consisted of four sections as follows:

- Section 1: Demographic and institutional data – Gender, age, Institution and Degree, Transfer from a Level 7 Degree.

- Section 2: Open Ended Question – participants were asked to state in their own words their understanding of SD.
- Section 3: Scaled Questions - this section was designed based on Azapagic et al [1]. Students were asked to rate their knowledge about SD principles, legislation, issues, SD tools and organizations that promote SD. The scale used was: **1: Never Heard of, 2: Heard but could not explain, 3: Have some Knowledge, 4: Know a lot.**
- Section 4: students were asked to rate the importance of SD at a personal and professional level.

The questionnaire was distributed both online and on paper in order to enhance the response rate and improve the results. In total, 143 fourth engineering students completed the questionnaire with a response rate of 54% across all disciplines. These included 48 Civil, 37 Mechanical, 26 Structural, 25 Building Services and 7 Chemical Engineers. It can be seen from Table 2 that most respondents came from one institution. The response from Institution 3 was very low making it difficult to make comparisons across institutions.

The data were analysed with SPSS 18. Raw data from the scaled sections were analysed in regard to frequency, average and also cross-tabulated with the key variables to test the significance of their impact. Open-ended responses were coded based on the three pillars of SD (Table 3).

Within Table 3 each class corresponds to specific key words related with each pillar that occurred in responses. Class 1 corresponded to answers that mentioned all three pillars of SD. Classes 2-7 show all the possible combinations of the pillars that might occur in other responses. Class 8 corresponds to answers that quoted Brundtland's definition. Class 9 was created for answers that showed no knowledge of SD.

Class 1 could be identified as a Relational class based on Carew's and Mitchell [6] taxonomy. Classes 2, 3 and 4 as Multi-structural; 5, 6, 7 and 8 as Uni-structural and 9 as pre-structural.

Classes	Key words
1. Economy-Environment-Society	**Economy**: long term planning, cost-payback analysis, development, growth, save money, economic **Environment**: environmental protection, limitations, eco friendly, impact, footprint, waste minimization, non and renewable resources **Society**: community, needs, social balance, equity.
2. Economy-Environment	
3. Economy-Society	
4. Environment-Society	
5. Economy	
6.Environment	
7.Society	
8. Brundtland Definition	
9. No evidence of knowledge	

Table 3: Classes used for coding open-ended responses.

The results are presented and discussed in the following section.

RESULTS

Aggregated results for all topics in the questionnaire indicated that engineering students' knowledge of SD is between **"Heard but could not explain"** and **"Have some knowledge"** with a corresponding overall average of 2.49. Although the overall average of this study is higher than Azapagic's 2.23 [1], further comparison showed that both studies illustrate the same overall pattern in engineering students' knowledge and understanding. When comparisons were drawn on topics common to both studies, it was found that averages for three sections fell closer to those found by Azapagic et al [1]. In one case, SD principles, the average fell below the average found in Azapagic's study (Figure 1).

The data shows that fourth year engineering students from the three Irish Higher Education institutions seem to be more knowledgeable in regard to sustainability issues with an average score of 2.86. Substantial knowledge gaps were identified in regard to SD principles and SD legislation, policy and standards. Figure 1 represents the average scores of this study and Azapagic's comparable average scores.

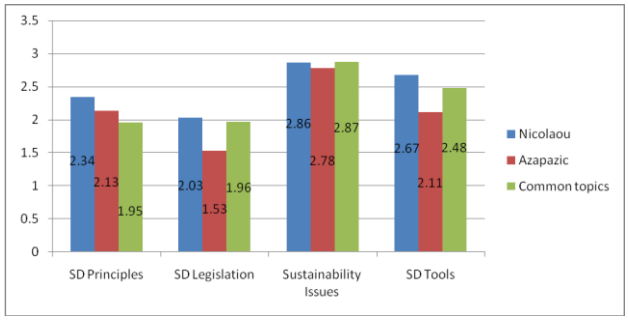


Figure 1: Section averages compared with Azapagic et al [1].

The scores show an inadequate level of knowledge and understanding of fourth year engineering students of SD in the Irish context.

SD principles

In relation to SD principles question, the overall average of 2.34 is just above "Heard but could not explain". Students rated their knowledge high in regard to the topics of "Engineering Ethics" with an average of 3.12 and "Minimizing the utilization of non-renewable resources" with an average of 3.09. Yet, considerable knowledge gaps were identified with principles regarding social issues. The latter seems to contradict with the high average score for "Engineering Ethics".

Student's ratings in regard to topics "Inter-generational equity", "Intra-generational equity" and "Social Inclusion" were below 2 which is just above "Never Heard of". More specific, "Inter-generational equity" had an average score of 1.48, "Intra-generational equity" 1.47 and "Social Inclusion" an average score of 1.88. Despite the higher overall average of this section, Inter and Intra generational had a lower average score than Azapagic's score (1.67). Moreover several other topics had a low score such as "Stakeholder Participation" (1.9) compared to Azapagic's (1.67); "Principle of Subsidiarity" (1.65) and the "Precautionary Principle" with an average of 1.58.

SD legislation, policy and standards

Turning to consider SD legislation, policy and standards it was found that three topics had a considerably higher average score than the section's overall average (2.03): "Kyoto Protocol" (2.94); "Ireland's Renewable Energy Targets" (3.01) and "Environmental Impact Assessment (EIA)" (2.92). Following the very low averages on social issues in the SD principles question, students' knowledge about legislation regarding social inclusion was significantly low. The topic "Aarhus Convention" had an average score of 1.19. Moreover, topics such as "Rio Declaration" and "Agenda 21" had low average scores of 1.42 and 1.43 respectively.

SD issues

The literature clearly states that engineering students tend to connect SD with environmental issues [1] [5] [6] [7] [8] [9]. Our data presents further evidence to support this view.

As previously stated, SD issues had the highest overall average score in this survey with a score of 2.86. Students gave a very high score to a number of environmental issues including "Climate Change" (3.49). Other topics in this section received an average score close to 3. However, relatively low scores were recorded for the topics "Ecosystems" at 2.47 and "Loss of Biodiversity" at 2.26. The topic "Loss of Biodiversity" was scored as the lowest also in Azapagic's study (2.21).

SD tools

In regard to SD tools section, students' responses showed an overall average score of 2.67, which was significantly higher than Azapagic's respondents (2.11). A higher score was given to the topic of "Recycling" (3.30). Average scores of 3.16 and 3.22 were recorded for the topics of "Renewable Energy Technologies" and "Use of renewable materials" respectively. Substantially lower than the overall average was the topic of "Tradable Permits" with an average score of 1.69 which it was also low in Azapagic's work with an average score of 1.82.

Organizations that promote SD

Students were also asked to rate their knowledge about several engineering bodies and international organizations that promote SD. The overall average of this section was 2.57 which lies between the statements of "Heard but could not explain" and "Have some knowledge". In this section, students had a good knowledge about "Engineers Ireland" with an average score of 3.39 and also about "Sustainable Energy Authority Ireland (SEAI)".

On the other hand a low average score of 1.73 was allocated to "Comhar, Sustainable Development Council" which is the National Development Council for Ireland. Comhar has published SD principles which Engineers Irelands have subscribed to.

A low average score of 1.97 was given to the "Intergovernmental Panel on Climate Change (IPCC)" topic.

Importance of SD

In Section 4, engineering students were asked to rate the importance of SD at the personal and professional level. 43.1% of the students rated SD as "Important" at a personal level, while 44.8% rated SD at a personal level as "Very Important". This corresponds to an average of 3.32 in a scale of 1-4. Azapagic et al [1] used the same scale to analyze this particular question which showed an average of 3.

On the other hand students' rating of the importance of SD at a professional level was higher that the personal level; 98.3% said that SD is either "Important" or "Very Important". More specifically, 72.4% rated SD as "Very Important" and 25.9% as "Important" at a professional level. This average 3.7 was higher, than Azapagic's average score of 3.3.

Key variables significance tests

Significance tests were performed using the Pearson Chi-square significance test in SPSS 18. Any chi-square smaller than 0.05 (<0.05) shows a significant impact of the testing key variable on the data. Age was the only key variable that had no impact on the survey data. On the other hand, on the aggregated 65 topics included in all sections, institution had a significant impact on 9 of them corresponding to 14%. This is not surprising given that the majority of respondents came from one institution.

Azapagic and her co-researchers [1] found that their results were not affected at all by key variables that had a connection with the participants' studies including their discipline.

However, chi-square tests in this study showed that degree was the primary variable impacting the results. Cross-tabulation of the degrees with the topics and chi-square performance showed that degree had a significant impact on the students' knowledge of 39 of 65 topics in the questionnaire, amounting to 60% of the topics surveyed. This suggests that engineering students' knowledge of SD is discipline-led.

Knowledge of nine of eighteen topics (50%) included in SD principles section; twelve of fourteen topics (85.7%) included in SD legislation section; seven of fifteen topics (46.7%) in SD tools; eight of twelve topics (66.7%) in SD issues and three of six topics (50%) in the organizations section were affected by the students' disciplines.

What seems to emerge from this analysis is that some disciplines are addressing discipline specific issues relatively well. What also emerges is that some issues such as Climate Change, Recycling and Deforestation score high independently of the discipline.

Another set of issues which include key social issues such as social equity and inclusion but also important legislation such as the Aarhus Convention score low regardless of discipline.

It is also worth noting that knowledge of the "Precautionary Principle" is scored low across all disciplines.

SD definition

Open-ended responses were coded based on the classification presented in Table 4. Figure 2 shows the results of the analysis. Based on the classification, engineering students define SD as a concept that is connected mainly with economic and environmental issues while failing to acknowledge the equal importance of the social aspect in SD.

Classes	1. Economy-Environment-Society
	2. Economy-Environment
	3. Economy-Society
	4. Environment-Society
	5. Economy
	6.Environment
	7.Society
	8. Brundtland Definition
	9. No evidence of knowledge

Table 4: Classes utilized for open-ended responses classification

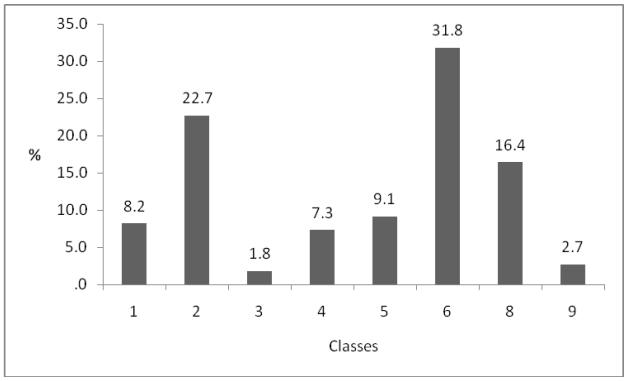


Figure 2: Classification of open-ended responses

Responses that define SD as a concept that consist of the three pillars, the economy, the environment and the society are classified as Class 1 which corresponds to a very low percentage of 8.2%.

A large group of participants (31.8%) described SD as strictly an environmental concept. In an attempt to aggregate all the classes that involve the environmental pillar, excluding Class 1, a large percentage of participants (61.8%) included the environment in their description of SD.

On the other hand, an aggregated percentage of the societal pillar is low at 9.1%, with Class 7 not having any responses.

Interestingly Brundtland's definition corresponds to 16.4 % of the responses which supports the fact that is a commonly used definition of SD.

All of the above show that fourth year engineering students have a very narrow understanding of SD which supports the findings from the scaled questions where results showed that students know very little about social issues.

An initial classification of these responses based on Carew's and Mitchell [6] taxonomy, showed that responses from this study are primarily uni-structural (Figure 3). Class 8 responses were classified as uni-structural in line with the approach taken by Carew's and Mitchell.

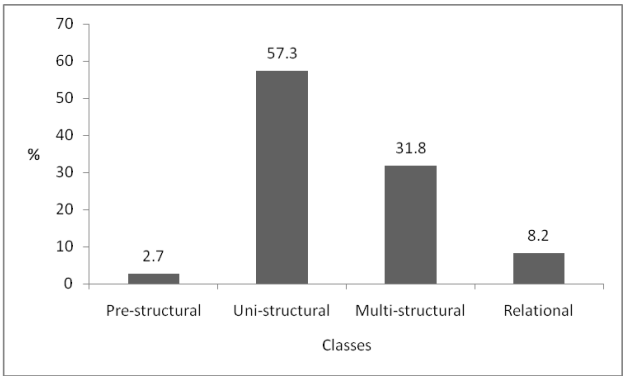


Figure 3: Classification of open-ended responses based on Carew and Mitchell's analytical framework

Classification of responses from this study, based on the SOLO taxonomy, follows a similar pattern as those from Carew's and Mitchell study. The majority were classified as uni-structural: Carew and Mitchel was 55.8% while ours was 57.3%.

While the proportion of Relational responses was lower in our study (8.2% as against 13.5%), there was a smaller proportion of Pre-structural answers (2.7% as against 9.6%). Both studies show that most students have a uni or multi-structural understanding of SD.

Figure 3 supports the argument rising from this study that students' understanding of the complexity of SD is very low. Students fail to acknowledge the inter-connectedness of the three pillars and tend to relate SD only with environmental issues.

CONCLUSIONS

The survey conducted for the purposes of this study showed that engineering students' knowledge in regard to SD topics is inadequate. It is concerning that respondents were fourth year engineering students who are just about to graduate.

The results follow the same pattern and identify the same knowledge gaps in engineering students' knowledge of SD as identified in Azapagic's et al study.

Substantial knowledge gaps are identified in regard to SD social issues, and SD legislation, policy and standards. Key SD issues such as "Loss of Biodiversity" and "Ecosystems" had a score significantly lower than the overall average which shows that engineering students also have significant knowledge gaps in environmental assessment and protection.

Several topics with a significantly high score such as "Recycling", "Climate Change" and the "Kyoto Protocol" have a high public profile. It is the case that these issues have achieved media coverage and are much discussed in political and public forums. It might be the case that students' knowledge is significantly impacted by the media coverage of these issues.

It is interesting to note that while students seem to know a lot about Climate Change, they have a poor knowledge of topics such as "Intergovernmental Panel on Climate Change (IPCC)" and "Tradable Permits". These topics relate more to the economic and political domain rather than to engineering.

Engineering students see the implementation of SD to be more of a professional requirement rather than an issue of personal commitment. They are more likely to say that it is very important for them as professionals. This suggests that the curriculum may not be generating a personal commitment to sustainability among engineering students and they might see sustainability as yet another constraint that engineers must grapple with in their engineering practice.

Fourth year engineering students' level of understanding of the complexity of SD is low. They directly connect SD with environmental issues while they neglect the social aspect of SD. They tend to define SD by referring to only one of the three pillars. Very few demonstrate Relational understanding of the concept while none had an Extended Abstract conception. It is the case thought that very small numbers have no knowledge at all.

This study has also showed engineering students' level of knowledge and understanding to be discipline-led.

SD is a multi-disciplinary concept that requires the involvement of the engineering sector. However, if engineers' competence is low as this study has shown, sustainability issues will not be effectively assessed by engineers. As Beder [14] has argued, "engineers of the future are professionals that understand SD and provide solutions that are appropriate in the three aspects of SD".

The present study generates new research questions such as the reasons why students have a narrow understanding of SD. The next stage in this project will seek to build on the data reported here.

We will be surveying first year engineering students with the same questionnaire. This will allow us to determine the degree to which pre-engineering experiences are affecting students' knowledge of SD.

We will also explore staff's understanding of SD. It might be the case that while academics might believe they are giving adequate coverage to SD issues these might be too discipline (and technology) specific. The result might be that students are not getting the general and broad education that they need to fully understand SD.

REFERENCES

[1] A. Azapagic, S. Perdan, D. Shallcross, How much do engineering students know about sustainable development? The findings of an international survey and possible implications for the engineering curriculum, **European Journal of Engineering Education**, Vol. 30, No. 1, 2005, pp 1-19

[2] N. Ashford, Major challenges to engineering education for sustainable development, **International Journal of Sustainability in Higher Education**, Vol. 5, No. 3, 2004, pp 239-250

[3] S Johnston, H. McGregor, E. Taylor, Practice-focused ethics in Australian engineering education, **European Journal of Engineering Education**, Vol. 25, No. 4, 2000, pp 315-324

[4] S. Johnston, Sustainability, Engineering and Australian Academe, **Society for Philosophy and Technology Quarterly Electronic Journal**, Vol. 2, No. 3/4. 1997, p 92

[5] J.R. Herkert, Sustainable development, engineering and multinational corporations, **Science and Engineering Ethics**, Vol. 4, No. 3, 1998, pp 333-346

[6] A.L. Carew, C.A. Mitchell, Characterizing undergraduate engineering students' understanding of sustainability, **European Journal of Engineering Education**, Vol. 27, No. 4, 2002, pp 349-361

[7] J. Segalas, D. Ferrer-Balas, K.F. Mulder, Conceptual maps: measuring learning processes of engineering students concerning sustainable development, **European Journal of Engineering Education**, Vol. 33, No. 3, 2008, pp 297-306

[8] F. Kagawa, Dissonance in students' perceptions of sustainable development and sustainability – Implications for curriculum change, **International Journal of Sustainability in Higher Education**, Vol. 8, No. 3, 2007, pp 317-338

[9] World Commission on Environment and Development (WCED), **Our Common Future**. Oxford University Press. 1987

[10] J.B. Biggs, K.F. Collis, **Evaluating the Quality of Learning: The SOLO Taxonomy**, New York: Academic Press, 1982

[11] Engineering Education for Sustainable Development Observatory (EESD), **Report 2008**, Available at https://www.upc.edu/eesd-observatory/why/reports/EESD_Observer_2008_.pdf/view

[12] Comhar, Sustainable Development Council**, Principles for Sustainable Development**, 1997, Available at http://www.comharsdc.ie/_files/S.D.Principles.pdf

[13] K. Mulder, **Sustainable Development for Engineers**, 2006, Greenleaf Publishing, UK

[14] S. Beder, Toward an environmentally conscious engineering graduate, **Australasian Journal of Engineering Education**, Vol. 7, No. 1, 1996, pp 39-45

Impact of Mathematics and Physics Grade Point Averages on the Overall GPA for Construction Management Undergraduate Students

Pramen P. SHRESTHA & Nirajan MANI
Civil and Environmental Engineering Department, Howard R. Hughes College of Engineering
University Of Nevada Las Vegas
Las Vegas, Nevada, 89154, USA

ABSTRACT

The performance of Construction Management (CM) students in their undergraduate studies is believed to be dependent on their mathematics and physics performance. The overall performance of students is measured by their overall Grade Point Average (GPA) for this study. Similarly, the performance of students in mathematics and physics are measured by the respective GPAs in these subjects. This study analyses the correlation of the mathematics and physics GPAs with the overall GPA. The study includes data of CM students who graduated between 1982 and 2010 from the University of Nevada, Las Vegas (UNLV). The research hypothesis for this study is that there is a significant correlation between the overall GPA with the mathematics and physics GPAs of CM undergraduate students. The statistical test results showed that there is a statistically significant correlation between the overall GPA with the mathematics and physics GPAs.

Keywords: Mathematics, Physics, Construction Management, Undergraduate Study, Grade Point Average.

1. INTRODUCTION

American Council for Construction Education (ACCE) requires that construction management undergraduate students in accredited programs take a minimum of 15 semester credit hours of mathematics and science courses. In addition to this, these students should take a minimum of 20 semester credit hours in construction science and a minimum of 20 semester hours in construction. A minimum aggregate of both construction science and construction combined of 50 semester credit hours requires as per ACCE.

It is believed that construction science courses require a significantly higher level of knowledge, skill, and ability in mathematics and physics than do the construction courses. Therefore, students must perform well in mathematics and physics courses to perform satisfactorily in construction science courses. It is also believed that students must perform equally well in construction science and construction courses to develop the level of knowledge, skills and abilities that are critical to becoming a successful professional in the construction industry. It is obvious that the overall grade point average (GPA) of the students directly depends upon all of these courses; in other words, the higher the grade in all courses, higher the overall GPA. The University of Nevada, Las Vegas (UNLV) Construction Management program places an emphasis on ensuring that students take the necessary mathematics and physics courses so that they are able to successfully apply these courses' fundamental concepts in construction science and construction courses.

The majority of the Construction Management (CM) undergraduate students at the University of Nevada, Las Vegas (UNLV) are having difficulty in passing the mathematics and physics courses required by the ACCE. This is impacting the graduation rate; also, students are having problems passing the construction science and construction courses at UNLV. It is assumed that the lack of knowledge in mathematics and physics is affecting the GPAs of these courses and ultimately affecting the overall GPA of these students. To determine whether there is a relationship between mathematics and physics GPAs with overall GPA, data was collected of CM students who passed these courses and graduated in last 10 years. The results of this study will shed some lights on this issue.

2. HYPOTHESES

Research Hypotheses

The variables considered for this study are the mathematics GPA, the physics GPA, and the overall GPA at graduation. Two research hypotheses for this study were formulated, as shown in Table 1.

Table 1. Research Hypotheses.

Hypothesis No.	Research Hypotheses (H_a)
H_a1	A higher GPA in mathematics will result in higher overall graduation GPA of Construction Management students.
H_a2	A higher GPA in physics will result in higher overall graduation GPA of Construction Management students.

Null Hypotheses

The above research hypotheses are converted to null hypotheses to conduct the statistical test, as described in Table 2. The statistical test hypothesizes that the correlation coefficient between these variables is not significantly different from zero. Mathematically, it can be expressed as:

$$\beta_1 = \beta_2 = 0$$

Table 2. Null Hypotheses.

Hypothesis No.	Null Hypotheses (H_o)
H_o1	There is no relationship between the mathematics GPA and the overall graduation GPA of the Construction Management students.
H_o2	There is no relationship between the physics GPA and the overall graduation GPA of the Construction Management students.

For the null hypothesis to be false, the p-value must be less than or equal to 0.05. Given that the null hypothesis is true, the p-value represents the probability of observing a test statistic that is at least as large as the one that is actually observed.

3. LITERATURE REVIEW

Several studies have been conducted to determine the factors that have significant effects on students' academic performance. These studies have found that student characteristics, teaching effectiveness, gender, academic classification, performance in prerequisite courses, and overall academic ability play significant roles in the performance of students [4, 6, 8, 9].

Choudhury (2001) conducted a research to determine the factors that affect the student's performance in Environmental Control Systems Courses at the undergraduate level in the Department of Construction Science at a large university in the south central region of the United States. Ten factors were considered for the analysis. The population collected for the study consisted of the 223 students who attended such courses during the Summer Semesters of 1997 and 1998, the Fall Semester of 1997, and the Spring Semester of 1998. The multiple regression analysis showed that such courses are not correlated with class size. Personal characteristic variables, such as gender and academic classification, are inversely related to student performance. The study found that the overall academic ability of a student is positively correlated with student performance. The F-test that was conducted was found to be statistically significant at the 0.0001 level. The R-square value of the model was found to be 0.37. From the multiple regression analysis, the following model was developed:

$$GRADE = 62.48 + 0.05 * CLSIZE + 0.95 * SEMESTER - 2.04 * SEX - 2.3 * LEVEL + 0.64 * INSPIRE + 0.24 * FEELING - 0.93 * CONTRIB + 0.21 * UNDSTAND + 0.5 * SATISFY + 5.36 * GPA \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (1)$$

A study by Meltzer (2002) was conducted to determine the relationship between mathematics preparation and conceptual learning gains in physics [1]. Students' conceptual knowledge was assessed by administering a diagnostic test on physics concepts during the first and last days of class. The authors collected the data of students enrolled in Fall 1997 and Spring 1998 physics courses of Southeastern Louisiana University (SLU) and Fall 1998 and Fall 1999 physics courses of Iowa State University (ISU). These physics courses covered electricity and magnetism, including direct current circuits. The sample size for these courses offered at SLU was 45 and 37, respectively. For the courses offered in ISU, the sample size was 59 and 78, respectively. The authors found that student's normalized learning gains were not significantly correlated with their pre-test scores on a physics concepts test. However, the normalized learning gain was significantly correlated with mathematics pretest scores in a sample of one LSU physics course and two ISU physics courses.

Sidiropoulos et al. (2005) in investigating the factors that contribute to student success, provided new ideas on how these factors could be used to offer adaptive learning in studying economics in a web based environment [3]. The authors collected data from 200 students enrolled in economics and finance classes in the department of Applied Informatics at the University of Macedonia, Greece. The data collected included family condition and education, sex, age, GPA, previous knowledge in mathematics, and previous knowledge about computers and the internet. This study showed that girls gained significant higher marks than boys in the economics module. The study also found that the father's education

was a strong indicator of student success with the module. Students having more educated fathers (M.Sc. or Ph.D.) gained better marks in the module than other students. Finally, the research showed that students with overall high GPAs scored high marks in economic lessons.

Choudhury (2002) conducted research to examine the effects of overall academic capability of students on their performance in this course [5]. The overall academic capability of a student was measured by GPA. The sample size for this study was 329, taken from a population of students who attended Environmental Control System courses in Fall, Spring, and Summer semesters from 1997 through 2001. A simple linear regression technique was used to predict the following model:

$$GRADE = 10.76 + 37.15 * GPA - 4.17 * GPASQ + 2.08 * MAJOR \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Where, GRADE = student performance in terms of numerical grade, GPA = grade point average of a student, GPASQ = quadratic term of GPA, MAJOR = academic major of a student.

The research found that overall academic capability and the major of a student both have statistically significant effects on the student performance in the Environmental Control Systems courses. The R-square value for the model was found to be 0.59. This model could be helpful to the instructors to formulate teaching strategies.

A study was conducted by Orth (2004) to identify variables that could serve as predictors of student retention and success in an undergraduate construction management program [2]. The independent variables in this study were high school rank, high school GPA, high school class size, the number of high school science courses, the number of high school mathematics courses, the SAT composite score, matriculation age, gender, race, and residence. Data was collected from 343 students enrolled from Fall 1992 to Fall 1997 in the Building Science Department at Purdue University. A logistic regression model was developed, and the relationship was tested at the 0.05 significance level. The Wild Chi-Square test showed that the high school GPA and the number of high school mathematics courses taken with regard to graduating the Construction Management (CM) program were found to be statistically significant at $\alpha = 0.05$. These results showed that the students who have taken significant number of mathematics courses in high school also have a high graduation rate in the CM program. This indicated that a solid mathematics education is vital for Construction Management students to graduate from this program.

Shrestha and Shields (2009) found that the GPA of undergraduate students in Construction Science course is directly correlated with the GPA of the mathematics courses taken at the university level. They found a significant correlation between the mathematics GPA and the final test score of the students in this course. However, no significant correlation was found between the physics GPA and the final test score of this course. A multiple regression model was created to predict the final test score of CM students in this course using mathematics and physics as input variables; it was found to be not significant [7].

4. METHODOLOGY

The methodology of this study consisted of five steps. The first step was to perform a literature review to determine whether any other studies have been conducted relating to these hypotheses: 1) better performance in mathematics results in a higher overall GPA for undergraduate CM students; and 2) better performance in physics results in a higher overall GPA for undergraduate CM students.

The second step of the study was to collect the mathematics and physics grades as well as the overall GPA data for undergraduate CM students. Then correlation tests were conducted between the mathematics and physics GPAs with regard to the overall GPA. If a significant correlation was found between the mathematics and physics GPAs with regard to the overall GPA, then the final step of the study was to develop a multiple regression model that could predict the overall GPA by using the mathematics and physics GPAs as input variables.

These five steps were followed by an exploration of the data collected and also a discussion of the implications of this research. At the end of this paper, conclusions and recommendations for future research areas will be discussed.

5. DATA DESCRIPTION

The data of CM students who graduated from UNLV between 1982 to 2010 were collected randomly. These data were obtained from the Office of Undergraduate Advising at UNLV. In the data analysis, only students who had taken mathematics and physics courses at the university level were considered. According to ACCE accreditation, CM students had to take at least two mathematics and two physics courses. The grade point average of the mathematics and physics courses was calculated for each student to use in the analysis. The total sample size for the analysis was 77. Both a linear correlation analysis and a multiple regression analysis

were conducted to determine the significant relationship between overall GPA with the mathematics and physics GPAs of the students.

6. RESULTS

Once the data was collected, it was entered into a Statistical Package for Social Sciences (SPSS) for processing. The descriptive data analysis shows that the mean for the overall grade, the mathematics grade, and the physics grade was 3.04, 2.86, and 2.69, respectively. The mean and median are very close to each other. The low standard deviation value showed that there is a small variation in the data set.

Variables (GPA)	No. of Sample	Descriptive Statistics		
		Mean	Median	Standard Deviation
Overall Grade	77	3.04	3.03	0.35
Mathematics Grade	77	2.86	2.85	0.50
Physics Grade	77	2.69	2.65	0.59

A Pearson Correlation test was conducted to determine the correlation coefficient between the overall GPA with mathematics and physics GPAs. Before conducting this test, the data were tested for normality. The histograms were plotted for these data, and showed that data for the overall, mathematics, and physics GPAs were normally distributed. Figure 1 shows the histograms of the overall, mathematics, and physics GPAs of CM students.

Table 3. Descriptive statistics of overall, mathematics, and physics grade point averages.

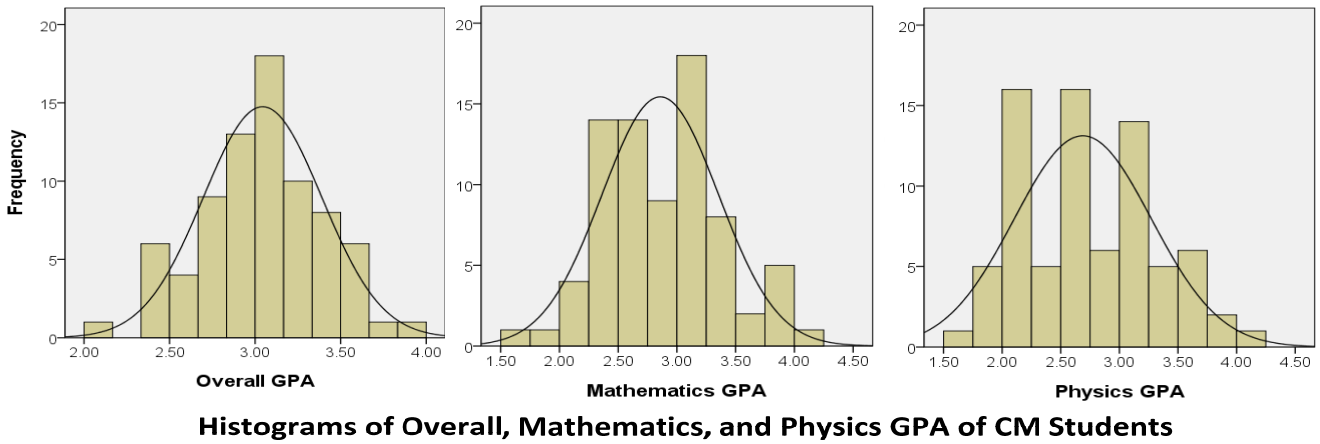


Fig. 1. Histogram of overall, mathematics, and physics GPAs.

Since histograms for mathematics GPA and Physics GPA curves do not show normal distribution, the Anderson-Darling test was conducted to check the normality of these variables. The null hypothesis of this test was that the data is normally distributed. If the p value was less than 0.05, then it would reject the null hypothesis. Table 4 shows the result of Anderson-Darling test, which indicated that all three variables are normally distributed.

Table 4. Anderson Darling Normality Test Results.

Variables	No. of Sample	p value	Remarks
Overall GPA	77	0.507	Normally distributed
Mathematics GPA	77	0.243	Normally distributed
Physics GPA	77	0.055	Normally distributed

The results of this analysis, presented in Table 5, indicate that there is a statistically significant relationship between

the overall GPA with the mathematics and physics GPAs. The correlation between the overall GPA and the mathematics GPA was found to be 0.497, and the correlation between the overall GPA and the physics GPA was found to be 0.503. These correlations were significant at alpha level 0.01. This also indicates that the relationships are positive, which indicates that the increase in mathematics and physics GPAs results in an increased overall GPA for CM students.

Table 5. Pearson Correlation test results.

Variables	No. of Sample	Pearson Correlation Coefficient	p value
Correlation with the Overall GPA			
Mathematics GPA	77	0.497**	<0.01
Physics GPA	77	0.503**	<0.01

** Significant at alpha level 0.01 (2-tailed)

Figure 2 shows the scatter plot of the overall GPA versus the mathematics GPA, and Figure 3 shows the scatter plot of the overall GPA versus the physics GPA. These graphs show that there is positive relationship between these variables.

The high correlation between these two variables suggests that a linear regression model can be developed to predict the overall GPA with the mathematics and physics GPAs of the students. Therefore, a multiple regression model was developed using the overall GPA as a dependent variable and the mathematics and physics GPAs as independent variables. The R-square value for the model was 0.247. Also, the model is statistically significant at alpha level 0.001. Equation (3) shows the regression model that predicts the overall GPA of CM students.

$$Overall\ GPA = 0.319\,x\,Mathematics\ GPA + 0.275\,x\,Physics\ GPA + 1.393 \quad ....(3)$$

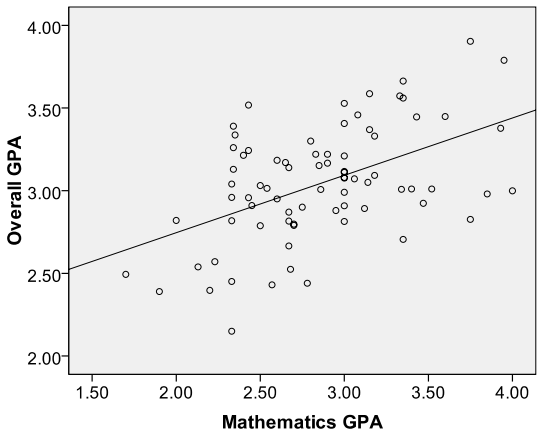


Fig. 2. Scatter plot of the overall GPA versus the mathematics GPA.

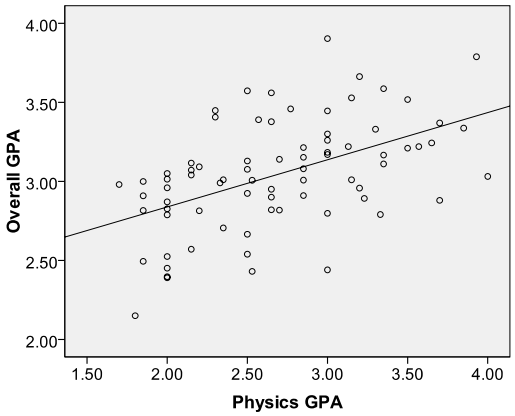


Fig. 3. Scatter plot of the overall GPA versus the physics GPA

7. CONCLUSIONS AND RECOMMENDATIONS

This study tested two major hypotheses. The first hypotheses, related to relationship between the overall GPA and the mathematics GPA, was proved to be true for this sample. The result showed that the better mathematics GPA is, the better the overall GPA of the CM students will be.

The second hypothesis states that there is a correlation between the overall GPA and the physics GPA, which also was found to be true for this sample. It indicates that the higher the GPA in physics, the higher the overall grade of CM students will be.

The CM students at UNLV take minimum of 15 credit hours of mathematics and physics courses before they start taking construction science courses. The number of credit hours taken in these two subjects is very low in compared to the credit hours taken in construction science and construction courses. Therefore the effect of mathematics and physics GPAs in the overall GPA is very small in comparison to the GPAs of other courses. However, this study results showed that the impact of the physics and mathematics courses on their overall GPA was very high and significant.

The findings of this research shed some light on how important it is for undergraduate students to perform well in mathematics and physics in order to achieve a higher overall GPA in the CM undergraduate program. It is recommended that further study be conducted to determine the impact of class size, instructor ability, types of instruction, gender, teaching effectiveness, and factors regarding undergraduate construction management subjects in order to determine the impact on the overall GPA.

8. REFERENCES

[1] C. E. Meltzer, "The Relationship between Mathematics Preparation and Conceptual Learning Gains in Physics: A Possible "hidden Variable" in Diagnostic Pretest Scores." **American Journal of Physics**, Vol. 70, No. 12, 2002, pp. 1259-1268.

[2] D. L. Orth, Identifying Predictor Variables of Student Success in a Construction Management Program. *ASC Proceedings of the 40th Annual Conference*, April, 2004.

[3] D. Sidiropoulos, A. Valasidou, & D. Bousiou-Makridou, Adaptive Web Based Learning and Factors Affecting Student Performance in Economics. *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications (EDMEDIA)*, pp. 4497-4501, June, 2005.

[4] I. Choudhury, "Correlates of student performance in environmental control systems courses at an undergraduate level." **Journal of Construction Education**, Vol. 6, No. 3, 2001, pp. 139-145.

[5] I. Choudhury, I. Predicting student performance using multiple regression. *Proceedings of the 2002 ASEE Gulf-Southwest Annual Conference,* March, 2002.

[6] J. Seymour, D. Goings, & A. Vincent, "Factors contribution to success in a microcomputer course." **Perceptual and Motor Skills**, Vol. 79, 1994, pp. 1338.

[7] P. P. Shrestha, & D. R. Shields, Correlating Student Performance in Fundamental of Construction Science Course with Mathematics and Physics Grade Point Average, *ASC Proceedings of the 45th Annual Conference*, pp. 33-41, April, 2009.

[8] R. J. Rose, C. W. Hall, & L. M. Bolen, and R. E. Webster, "Locus of control and college students' approaches to learning." **Psychological Reports**, Vol. 79, 1996, pp. 163-171.

[9] S. L. Coleman & A.J. Gotch, "Spatial Perception Skills of Chemistry Students." **Journal of Chemical Education Research**, Vol. 75, No. 2, 1998, pp. 206-209.

Affordable Platforms for High Performance Computing and Computational Science Education

Yili TSENG
Department of Electronics, Computer, and Information Technology
North Carolina A & T State University
Greensboro, NC 27411, U.S.A.

ABSTRACT

Traditionally, engineering and science disciplines have relied on theory and experiments as tools to perform research. With the introduction of computer and software, computer simulation based on mathematical modeling gradually becomes an important tool. The mechanism evolves into computational science. Computational science has become a tool as important as theory and experiments to all engineering and science disciplines nowadays. In most cases, it is adopted more often than experiments because it is more economic, less time-consuming, and able to explore infeasible situations. Computational science depends on high performance computers which refer to parallel computers. Parallel programs have to be written to take advantage of all processors in a high performance computer. Since 2003, the speed of uniprocessors can hardly be pushed because of energy-consumption and heat-dissipation problems. That is why the major processor manufacturers like Intel and AMD introduce multi-core processors instead of faster uniprocessors since then. Likewise, parallel programs have to be executed to take advantage of all cores of a multi-core processor. Both phenomena push the need for high performance computing education. However, the largest barrier to high performance computing and computational science education is the high cost of high performance computers. The author presents two low-cost platforms, namely commodity clusters and PCs equipped with GPUs, in this paper so that any institution can select the option which fits their affordability.

Keywords: Computational Science, High Performance Computing, Computing Education, Commodity Cluster, Parallel Programming, GPU Programming

1. INTRODUCTION

Traditionally, engineering and science disciplines have relied on theory and experiments as tools to perform research to explore new knowledge. With the introduction of computer hardware and software, computer simulation based on mathematical modeling gradually becomes an important tool. The mechanism evolves into computational science after high performance computers are mature and commercially available. Computational science has become a tool as important as theory and experiments to all engineering and science disciplines. In most cases, it is adopted more often than experiments because it is more economic, less time-consuming, and able to explore infeasible situations. Hence, computational science education should be included by engineering education.

Computational science depends on high performance computers. High performance computers refer to parallel computers, namely computers equipped with multiple processors. Although parallel computers are capable to execute sequential programs, but only one processor is being utilized. Parallel programs have to be written and executed to take advantage of all processors in a high performance computer. Therefore, at least introductory high performance computing and parallel programming courses should be covered by engineering education as well. Another rationale which makes both courses imperative for engineering education is described as follows. Since 2003, the speed of uniprocessors can hardly be pushed because of energy-consumption and heat-dissipation problems. That is why the major processor manufacturers like Intel and AMD introduce multi-core processors instead of faster uniprocessor since then. Just like multiprocessor computers, parallel programs have to be executed to take advantage of all cores of a multi-core processor. Both phenomenons push the need for high performance computing education.

However, the largest barrier to high performance computing and computational science education is the high cost of high performance computers. Most institutions cannot afford expensive parallel computers. Even if a university owns few high performance computers, they are reserved for research and would not be used for teaching and learning. That is the case at author's institution. In order to find affordable

platforms for his high performance computing classes, the author explored and managed to discover two options of affordable platforms suitable for high performance computing classes. Definitely they are suitable for computational science classes as well. The first platform is commodity clusters consisting of inexpensive personal computers (PCs) and network switches. They even can be built with retired PCs. The second platform is PCs equipped with Graphic Processing Unit (GPU). Both approaches provide low-cost solutions for all institutions to offer their high-performance computing and computational science education. They will be presented in this paper so that any institution can select the option which fits their affordability and built them with the minimum costs.

2. COMMODITY CLUSTERS

C/C++ and FORTRAN has dominated the numerical methods field, a key part of computational science, for decades and numerous of programs were coded in both languages. It is beneficial to stick with both languages and reuse exiting codes and libraries. Message Passing Interface (MPI) is a standard developed for parallel libraries supporting C/C++ and FORTRAN. [25][26] Lots of libraries are implemented after MPI for different platforms. MPI programs are portable among those platforms. Students learning MPI programming from small-scale learning-purpose machines can easily adapt to large-scale high performance computers. Hence, the first approach is to execute MPI programs on commodity clusters.

Thanks to the contribution of open-source software developers, MPI libraries have been successfully ported to inexpensive PC platform. Along with other free open-source operating systems and applications for PC, they can make PC's networked by low-cost switches a commodity cluster, an affordable platform for high performance computing and computational science education. With clusters built with the retired PC's and free software, any institution can own its platforms with minimal cost and start high performance computing and computational science education. Although clusters built with retired PCs do not have sufficient computing power to execute decent parallel applications, they do exhibit all characteristic of parallel processing and can execute qualitative experiments. If an institution owns sufficient funding, it can acquire high-end PC's and construct a cluster which has decent computing power to execute serious parallel programs for research.

While an affordable cluster can be built with the free open-source software and retired PC's, it cannot work practically without some vital configurations. Several books have been written about building a cluster with Linux. [2][3][4][5][6] However, all of them fail to point out the vital configurations required to make the cluster work correctly. The practical issues [11] in building an inexpensive cluster are addressed in the following subsections respectively.

Hardware Requirements

Any PC with 128B RAM or more and an Ethernet network interface card (NIC) can work as a node. PCs with Pentium III 500MHz CPUs work smoothly at author's institution. All nodes have to be connected with a network hub or switch. Generic Ethernet NICs and switches can be acquired with very little cost. The logical layout is shown in Figure 1 and the physical configuration is displayed in Figure 2. KVM (keyboard, video, and mouse) switches shown in Fig. 3 can be installed to save space and devices.

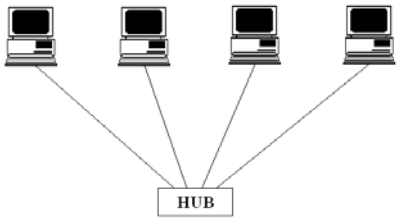


Fig. 1, Logical layout of the cluster

Fig. 2, Physical configuration of a 4-node cluster

Fig. 3, KVM (keyboard, video, mouse) switch

Operating System and Software Packages

Among all operating systems, only Linux can be acquired for free. Also, Linux operating systems come with plenty of hardware drivers which cover almost all legacy and new hardware, that further makes it the ideal OS for commodity clusters. Although several implementations of open source Linux operating systems are available, not all of them work well with the MPI library. After extensive experiments, the author chose and installed CentOS Linux which is a clone of commercial Red Hat Enterprise Linux [9] and downloadable at www.centos.org. The following actions should be done for all nodes in the cluster. Firewall and SELinux should be turned off during installation as they cause difficulty for communications among nodes which are required for executing MPI programs. Fortunately, security is not a concern as long as the cluster is not connected with other networks. The editor, gcc C++ and FORTRAN compilers under "Development Tools" have to be installed to as they are required for MPI programming process. One node should be selected as the server node and the following software should be checked and installed on the node during installation: Network Information Service (NIS) server under "Network Server" and Network File System (NFS) server configuration tool under "Server Configuration Tool." After Linux is installed, networking should be correctly configured so that all nodes of the cluster can communicate among one another. In general, one user account with the same name needs to be created on each node because a parallel program is dispatched to each node under the same user account. NIS which is addressed in later subsections can take care of this and other issues.

MPI library

The next major step is to install the MPI library. Again, there are several implementations of MPI, such as MPICH, LAM/MPI, Open MPI, etc. Nevertheless, only Open MPI is still under active development and growing more powerful. Therefore, Open MPI is the best choice. The steps to install Open MPI are quite straightforward. They are described as follows. First, download the latest version of the library from Open MPI's website, www.openmpi.org. Log into the root account to install it. Copy the compressed file to the /tmp directory. Uncompress the file with the command line [8] or double click to invoke GUI uncompress software:

shell$ gunzip -c openmpi-1.4.3.tar.gz | tar xf –

Then change to the directory openmpi-1.4.3. Configure and make Open MPI with the commands below. [7] Replace the directory after prefix option if you want to install into another directory.

shell$ cd openmpi-1.4.3

shell$ ./configure - -prefix=/usr/local

shell$ make all install

The make process may take up to one hour on an old PC.

NFS and NIS

Before we run a parallel program on our cluster, we need to dispatch a copy of the program's executable file onto every node under the same account. Manually copying the executable to all nodes is impractical. Network File System (NFS) is the solution to help us to complete this mission. With NFS, the program's executables only need to be saved into the shared directory of the NFS and a copy of the program is automatically copied to all other nodes. While NFS can solve the copying problem, there are some potential security problems. Each user has access to the shared directory, meaning any user can run, correct, save, and delete others' programs. To remedy this problem, Network Information System (NIS) was used to create accounts on the server node. All users can login from any node and manage their own shared directory. You have to login as root to perform all following setups and reboot all nodes to take effect. Do not reboot any client node until the server node completes its boot-up process, otherwise other nodes cannot read the correct configuration information from the server and perform normally.

Set Up the NFS Server [9]

1) From the *NFS Server Configuration* window, click *File → Add Share*. The *Add NFS Share* window appears. In the *Add NFS Share* window Basic tab, type the following information:
    - Directory – Type the name of the directory you want to share. Type "/home" which is the parent directory to all user directories.
    - Host(s) – Enter one or more host names to indicate which hosts can access the shared directory. Type "*" to let all nodes access NFS server.
    - Basic permissions – Click *Read/Write* to let remote computers mount the shared directory with read/write access.
2) To permanently turn on the NFS service, type:

shell$ chkconfig nfs on
shell$ chkconfig nfslock on

Set Up the NFS Client [9]

To set up an NFS file system to mount automatically each time you start your Linux system, you need to add an entry for that NFS file system to the /etc/fstab file. The /etc/fstab file contains information about all different kinds of mounted (and available to be mounted) file systems for your Linux system.
The format for adding an NFS file system to your local system is the following:
host:directory mountpoint options 0 0

The first item (host:directory) identifies the NFS server computer and shared directory. Mountpoint is the local mount point on which the NFS directory is mounted, followed by the file system type (nfs). Any options related to the mount appear next in a comma separated list. For our system, we add the following NFS entries to /etc/fstab:

kingtiger1:/home /home nfs rsize=8192,wsize=8192 0

Set Up the NIS Client [9][10]

1) Defining an NIS domain name:
   Our NIS domain name is kingtiger, we can set it by typing the following as the root user from the shell:

   shell$ domainname kingtiger

   To make the NIS domain name permanently, you need to have the domainname command run automatically each time your system boots. We can do it by adding the command line to a run-level script that runs before the ypbind daemon is started. We edited the /etc/init.d/network file and added the following lines just after the first set of comment lines.

   domainname kingtiger

2) Setting up the /etc/yp.conf file:
   We had an NIS domain called kingtiger and a server called kingtiger1, we should have the following entries in our /etc/yp.conf file:

   domain kingtiger server kingtiger1
   domain kingtiger broadcast
   ypserver kingtiger1

3) Configuring NIS client daemons:
   We need set up an existing run-level script called ypbind to start automatically at boot time. To do this, you can run the following command:

   shell$ chkconfig ypbind on

4) Using NIS maps:
   For the information being distributed by the NIS server to be used by the NIS client, you must configure the /etc/nsswitch.conf file to include nis in the search path for each file you want to use. In most cases, the local files are checked first (files), followed by nis. The following are examples of how some entries should be changed:

   passwd: files nis
   shadow: files nis
   group: files nis
   hosts: fiels nis dns

   As soon as the /etc/nsswitch file is changed, the data from NIS maps are accessible. No need to restart the NIS service.

Set Up the NIS Server [10][9]:

1) To configure your Linux system as an NIS server, you should first configure it as an NIS client and reboot the system.
2) Creating NIS maps:
   To create NIS maps so that your Linux system can be an NIS server, start from the /var/yp directory from a Terminal window as root user. In that directory, a Makefile enables you to configure which files are being shared with NIS. All default configurations in Makefile are ok for our purposes, so we don't need change them.
3) Configuring access to maps:
   In the /etc/ypserv.conf file, you can define rules regarding which client host computers have access to which maps. For our purposes we just need add the following line into /etc/ypserv.conf to allow all hosts access to all maps:

   * : * : * : none

4) Configuring NIS serve daemons:
   We can use the following chkconfig command to set ypserv and yppasswdd scripts to start automatically at boot time.

   shell$ chkconfig ypserv on
   shell$ chkconfig yppasswdd on

5) Updating the NIS maps:
   If you modify the sources for NIS maps (for example if you create a new user by adding the account to the passwd file), you need to regenerate the NIS maps. This is done by a simple

   make –C /var/yp

   This command will check which sources have changed, creates the maps new and tell ypserv that the maps have changed.

Disabling Password Authentication

As Open MPI is configured by default to use ssh (secured shell) to dispatch parallel tasks, it is ssh that asks for password to authenticate the connection. Extra steps below will prevent ssh from requesting passwords [7]. Because the measure should only work for the user account which intends to run MPI applications, log into the user account instead of root account of server node, to configure. First, generate the private and public key for the user account by executing:

shell$ ssh-keygen -t dsa

Next, ssh to all other nodes respectively by typing:

shell$ ssh host_name

That will generate the hidden .ssh directory on each node with the necessary attribute. Remember to go back to the server node by typing "exit" to finish the remaining procedures. Change into the .ssh directory under the user account and do the following [7].

shell$ cd /home/user/.ssh
shell$ cp id_dsa.pub authorized_keys

Next, repeat the following steps for all other nodes in the cluster. Replace "user" with account names which you want to disable requesting passwords.

shell$ scp authorized_keys host_name:/home/user/.ssh
shell$ scp id_dsa host_name:/home/user/.ssh
shell$ scp id_dsa.pub host_name:/home/user/.ssh

With these procedures done, all private, public, and authorized keys are duplicated on each node. They will be used for authentication for all future connections without passwords being requested from other nodes. Now the MPI applications can be executed on multiple nodes of this cluster without being asked for passwords.

Running MPI applications
If you do not add the path to your login profile, you should add the path in front of mpicc and mpicrun [7]. The path is /usr/local/bin in this example.

On Single Node: You should log into the user account to compile and execute MPI programs. MPI applications can be compiled with the command [7]:

shell$ mpicc mpi_program.c –o mpi_program

MPI applications should be executed with the command below [7]. Replace the number in the np option (number of process) with any value you want.

shell$ mpirun -np 4 mpi_program

On Multiple Nodes: In order to run MPI applications on multiple nodes, you need to specify the hostnames of all participating nodes in a file, my_hostfile in the example, and execute with the option hostfile [7].

shell$ mpirun - -hostfile my_host -np 8 mpi_program

3. GPU-EQUIPPED PERSONAL COMPUTERS

Computer Unified Device Architecture (CUDA) is an architecture developed by nVidia for its GPUs. [27][28][29] CUDA adopts many-core approach which has numerous much smaller cores than the cores of a multi-core CPU. The massive GPU cores are optimized for floating-point calculations while the CPU cores are optimized for sequential code execution. [28] Computational science applications rely heavily on floating-point calculations. CUDA can place much more cores in a chip than multi-core CPUs. That makes it more powerful than multi-core CPUs in floating-point calculations as more cores are dedicated to the purpose. In 2009, the ratio for peak floating-point calculation throughput between many-core GPU and multi-core CPU is about 10 to 1. [28] GPUs' design philosophy also makes them cheaper than multi-core CPUs. Table. 1 lists the numbers of GPUs and costs of four different level graphic cards. The prices are found on www.pricewatch.com as of June 2011. Apparently, the cost/performance ratio of GPUs is much better than that of multi-core CPUs. The programming language is CUDA C which is similar to MPI and OpenMP. Lots of computational scientists and researchers are adopting GPU programming. Therefore, GPU computing is very promising for computational science and high performance computing and should be offered by institutions.

Model	Number of GPU	Price
GeForce 210	16	$35
GeForce GT 430	96	$68
GeForce GTX 460	336	$155
GeForce GTX 580	512	$452

Table. 1 Features and Costs of GPU

With the low-costs of GPUs, PCs equipped with GPU graphic cards are the second approach affordable for any institution. The CUDA library and compiler is provided by nVidia for free for Linux, Windows, and MacOS. All software required for GPU programming is free if Linux platform is adopted.

Hardware Requirement
Any graphic card equipped with nVidia GPU can be utilized. GPU Device drivers are required and can be downloaded from www.nvidia.com. [27]

Software Requirement
The software required for GPU programming is C compiler for GPU and C compiler for CPU. The GPU compiler is contained in CUDA Development Toolkit which is freely downloadable at nVidia's web site. [27] The GNU C compiler coming with CentOS can be used as the compiler for CPU.

4. CONCLUSIONS

As the development of processors has shifted to multi-core approach, parallel programming becomes imperative to effectively utilize the extra cores of modern CPUs. Computational science has evolved to become the most important tools for research in many engineering and science fields. In turn, computational science depends on high performance computers which are in fact parallel computers. Parallel programming is required for computational science. Hence, high performance computing and computational science education is important for all engineering and science majors. Nonetheless, the barrier for most institutions to offer both disciplines is the high cost of high performance computers as not many institutions can afford the costly parallel computers. Due to the contribution of open-source software developers, MPI parallel library has been ported to inexpensive PC platform. That opens the door to affordable high performance computing. A commodity cluster can even be built with retired PCs which any institution must have. Nevertheless, all vital configurations to make a cluster work smoothly are scattered everywhere. The author presented all vital configurations in this paper to enable institutions with few or no resources at all to build commodity clusters with minimal costs to offer high performance and computational science courses.

The author also introduced GPU programming as another option affordable platform. Owing to the amazing performance in terms of cost, the GPU programming introduced by nVidia is attracting lots of attentions in high performance computing field. Like the commodity clusters, the software required by GPU programming can be acquired for free. Combined with the low cost of graphic cards with less GPUs, the platform for GPU programming can be afforded by any institution.

5. REFERENCES

[1] Grama, Ananth et al., **Introduction to Parallel Computing**, Addison-Wesley, 2003

[2] Sloan, Joseph D., **High Performance Linux Clusters**, O'Reilly, 2005

[3] Bookman, Charles, **Linux Clustering**, New Riders, 2003

[4] Vrenios, Alex, **Linux Cluster Architecture**, Sams, 2002

[5] Lucke, Robert, **Building Clustered Linux Systems**, 2005

[6] Gropp, William et al., **Beowulf Cluster Computing with Linux**, 2nd Ed., MIT Press, 2003

[7] www.openmpi.org FAQ

[8] Pacheco, Peter S., **Parallel Programming with MPI**, Morgan Kaufmann Publishers, Inc. 1997

[9] Sobell, Mark, **A Practical Guide to Red Hat Linux**, Third Edition, Prentice-Hall, 2006

[10] http://www.linux-nis.org/nis-howto/

[11] Tseng, Y. and Hargrove, C, "Construction of Parallel Machines for Engineering Courses with Retired Personal Computers", **ASEE Southeastern Section Annual Conference**, Memphis, TN, 04/2008

[12] Hughes, C. and Hughes, T., **Professional Multicore Programming**, Wiley, 2008

[13] Hughes, C. and Hughes, T., **Parallel and Distributed Programming Using C++**, Addison-Wesley, 2004

[14] Reinders, James, **Intel Threading Building Blocks**, O'Reilly, 2007

[15] Nichols, B. et al., **Pthreads Programming**, O'Reilly, 1996

[16] Welling, A., **Concurrent and Real-Time Programming in Java**, Wiley, 2004

[17] Breshears, C., **The Art of Concurrency**, O'Reilly, 2009

[18] Parhami, B., **Introduction to Parallel Processing**, Plenum, 1999

[19] Buyya, R., **High Performance Cluster Computing**, Vol. 2, Prentice-Hall, 1999

[20] Magee, J. and Kramer, J., **Concurrency**, Wiley, 2006

[21] Yang, L. and Guo, M., **High-Performance Computing**, Wiley, 2006

[22] Kepner, J., **Parallel MATLAB**, SIAM, 2009

[23] Arora, G. et al., **Microsoft C# Professional Projects**, Premier, 2002

[24] http://software.intel.com/en-us/blogs/2006/10/19/why-windows-threads-are-better-than-posix-threads/

[25] Gropp, W. et al., **Using MPI**, MIT Press, 1999

[26] Gropp, W. et al., **Using MPI-2**, MIT Press, 1999

[27] Sanders, J. and Kandrot, E., **CUDA by Example**, Addison-Wesley, 2011,

[28] Kirk, D. and Hwu, W., **Programming Massively Parallel Processors**, Morgan Kaufmann, 2010

[29] Hwu, W., **GPU Computing Gems Emerald Edition**, Morgan Kaufmann, 2011

The Roles of Grain Boundary Energy Anisotropy and Second–Phase Particles on Grain Growth in Polycrystalline Materials

Mohsen Asle Zaeem [*], **Haitham El Kadiri, Mark F. Horstemeyer, Paul T. Wang**

Center for Advanced Vehicular Systems, Mississippi State University

Starkville, MS 39759, USA

[*] mohsen@cavs.msstate.edu, +1 662-325-0126

ABSTRACT

A phase-field model was developed to investigate the concurrent effects of grain boundary energy anisotropy and second-phase particles on grain growth in polycrystalline materials. The phase-field model was developed based on the evolution of non-conserved phase-field variables according to the time-dependent Ginzburg-Landau (TDGL) equation. The Read-Shockley and modified Read-Shockley models for cubic crystals were considered to include anisotropic grain boundary energies for low and high misorientation angels between adjacent grains. Systems without particles reach a steady state grain growth rate. The presence of particles significantly alters the microstructures during grain growth. Results show that for systems with particles, the critical average grain size to stop grain growth depends on both the volume fraction and size of particles and also on the grain boundary energy anisotropy.

Keywords: Phase-field model, Grain growth, Anisotropic grain boundary energy, Second-phase particles, Finite element.

1. INTRODUCTION

Grain growth is one of the most common microstructural evolutions resulting from different materials processing [1]. Size and shape of grain microstructures determine the mechanical and material properties of alloys [2], and conducting research in this area to predict and control the microstructures at different thermal and mechanical conditions is necessary for design optimization of engineering structures. There have been several numerical studies in this area based on different computational approaches such as Monte Carlo [3,4], and phase field [5,6] methods. Most of the current models consider only a single phase for the polycrystalline material and do not include the effects of second-phase particles. Also the effect of anisotropic grain boundary energy which can significantly change the kinetics of the microstructural evolution has not been included in the numerical models of grain growth. In this research, a phase-field – finite element model is used to simulate the microstructural evolutions during grain growth considering the simultaneous effects of anisotropic grain boundary energy and second-phase particles on the growth rate.

2. FORMAULATION

The polycrystalline microstructure can be described by many non-conserved phase-field variables or order parameters ($\eta_i(\mathbf{r},t), i=1,2,....,n$, where n is number of different orientations) [5]. These phase-field variables are continuous functions in time and space dimensions and represent different orientations for different grains (see Fig. 1). The value of the field variables varies between 0 and 1; for example for grains having the *q*th orientation: $\eta_{i=q}(\mathbf{r},t)=1$ and $\eta_{i\neq q}(\mathbf{r},t)=0$ and this transition from 0 to 1 in the grain boundaries is smooth. The total grain boundary energy of a microstructure is a function of these phase-field variables and their gradients:

$$F=\int_V \left[\ f(\eta_1,\eta_2,\ldots,\eta_n)+\sum_{i=1}^{n}\frac{\kappa_i}{2}(\nabla\eta_i)^2\ \right]dV, \qquad (1)$$

where κ_i are the gradient energy coefficients. The local free energy density, f, has this form:

$$f(\eta_1,\eta_2,\ldots,\eta_n)=\sum_{i=1}^{n}\left(-\frac{\alpha}{2}(\eta_i)^2+\frac{\beta}{4}(\eta_i)^4\right)$$
$$+\gamma\sum_{i=1}^{n}\sum_{j\neq i}^{n}\eta_i^2\eta_j^2+\varepsilon\Phi^2\sum_{i=1}^{n}\eta_i^2, \qquad (2)$$

α, β, γ, and ε are phenomenological parameters, and $\Phi=1$ inside a particle and $\Phi=0$ in the matrix. Here *n* is the number of different orientations for grains while no orientations are considered for particles. When $\Phi=1$, f has one minimum at all η_i equal to 0. In our simulations, the parameters of the model are considered

to be constants similar to original model by Fan and Chen [5]: $\alpha = \beta = \gamma = 1$ ($\varepsilon = \gamma$).

The evolution equations of non-conserved order parameters can be obtained from time-dependent Ginzburg-Landau (TDGL) [7] equation given by the following,

$$\frac{\partial \eta_i(\mathbf{r},t)}{\partial t} = -L_i \frac{\delta F}{\delta \eta_i(\mathbf{r},t)}, \quad i = 1,2,\cdots,n, \tag{3}$$

where L_i are related to the grain boundary mobility (in this paper, $L_i = 1$ similar to Fan and Chen model [5]).

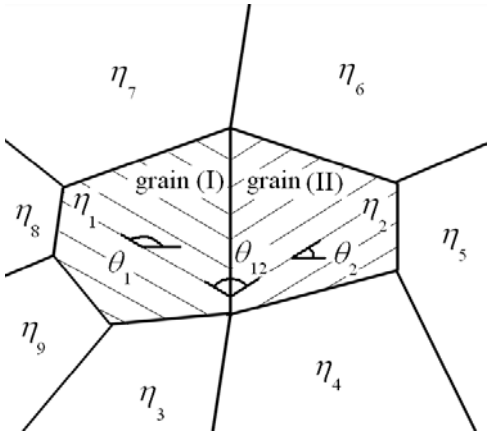


Fig. 1. Schematic of various crystal orientations using different orientation field variables for each grain. The solid-lines are grain boundaries.

The relationship between gradient energy coefficient and grain boundary energy, $E(\theta_{ij})$, [8]:

$$\kappa_i = \bar{\kappa}_j E^2(\theta_{ij}), \quad \theta_{ij} = \theta_i - \theta_j, \tag{4}$$

where θ_i is the orientation angle of a grain with respect to a reference angle, θ_{ij} is the angle between grains *i* and *j*, and $\bar{\kappa}_j$ is considered to be a constant (in this paper, $\bar{\kappa}_j = 2$, similar to the Fan and Chen model [5]). In our model to calculate θ_{ij} for each grain boundary between two adjacent grains, we assumed a random orientation angle ($\theta_i = [0 \quad \pi]$) corresponding to each order parameter (see Fig. 1).

The grain boundary energy considering low misorientation angles between adjacent grains ($\theta_{ij} < 20^\circ$) is presented by Read and Shockley (RS) [9]. The conventional form of RS model is [10]:

$$E(\theta_{ij}) = E_0 \frac{\theta_{ij}}{\theta_m}\left(1 - \ln\left|\frac{\theta_{ij}}{\theta_m}\right|\right), \quad \theta_{ij} < \theta_m, \tag{5}$$

where E_0 is a material constant and is proportional to the total density of dislocations (in our simulations $E_0 = 1$), and θ_m is the maximum misorientation angle for RS model ($\theta_m = 20^\circ$). For $\theta_{ij} \geq 20^\circ$, $E(\theta_{ij})/E_0 = 1$, same as isotropic cases. For large misorientation angles, modified RS (MRS), is developed by Wolf [11]:

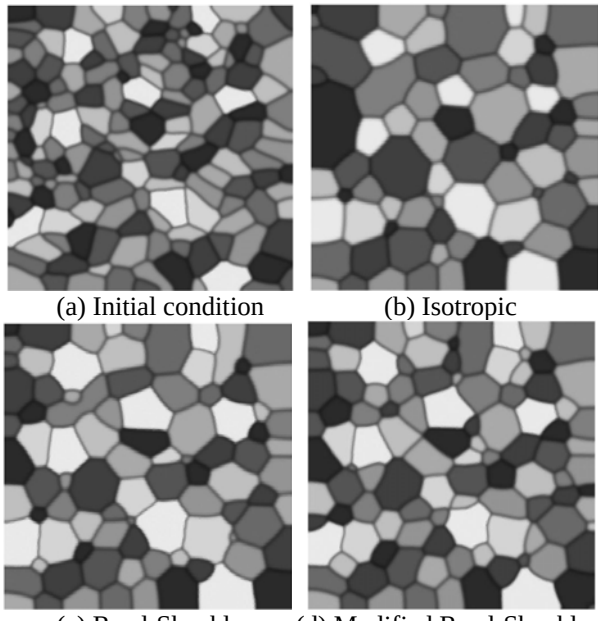
$$E(\theta_{ij}) = E_0 \sin(2\theta_{ij})\left(1 - r \ln\left|\sin(2\theta_{ij})\right|\right), \tag{6}$$

where *r* is a constant and in our simulations $r = 0.683$ to make $\left(E(\theta_{ij})/E_0\right)_{\max} = 1$. More details about this model can be found at [12].

3. RESULTS

Eq. (3) was solved in a 400×400 square domain using the finite element software, COMSOL multiphysics [13]. A uniform mesh consists of 22000 triangular elements with quadratic interpolation functions are used to discretize the domain for all simulations. Time step size of iterations is $\Delta t = 0.2$. To simulate the polycrystalline material properly, 36 order parameters are used [5]. For all simulations, a same initial grains configuration is chosen to be a Voronoi tessellation diagram with 36 random initial orientation angles corresponding to order parameters, Fig. 2 (a). For cases including particles, particles are randomly distributed in the grains system where V_f and r_p are the volume fraction and size of particles, respectively.

In Fig. 2, microstructural evolution at $t = 2000$ is shown for the isotropic, Read-Shockley and modified Read-Shockley models of grain boundary energy. This figure shows that including the effects of anisotropic grain boundary energy significantly alter the microstructures.



(a) Initial condition (b) Isotropic

(c) Read-Shockley (d) Modified Read-Shockley

Fig. 2. Microstructural evolution. In (b)-(d) $t = 2000$.

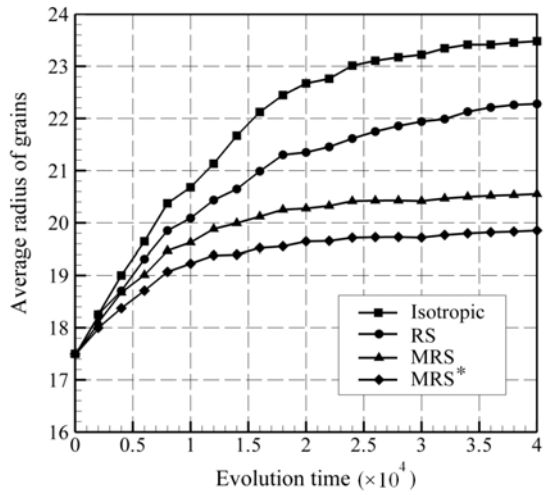


Fig. 3. Average grain radius versus time of evolution for systems with inert particles: $r_p = 2$, $V_f = 0.015$; $^*V_f = 0.03$.

The presence of second-phase particles slows or freezes the grain growth process and results in different microstructural configuration usually with finer grain sizes that can be accountable for different materials properties. Fig.3 shows the average grain radius versus time of evolution in systems with inert particles. From Fig. 3, it is evident that including anisotropic grain boundary energy significantly affects the grain growth morphology. The critical average grain size to stop the grain growth depends not only on volume fraction and size of particles, as predicted by other models, but also on the grain boundary energy anisotropy.

4. CONCLUSIONS

A phase-field model was developed to study the simultaneous effects of anisotropic grain boundary energy and second-phase inert particles on grain growth kinetics. Including the effects of anisotropic grain boundary energies decreases the grain growth rate, and this decrease is more significant when the effect of high misorientation angles is considered. Results showed that the critical average grain radius to stop the grain growth is a function of the volume fraction and size of second-phase particles and also significantly depends on the grain boundary energy anisotropy for systems with low volume fraction of particles.

5. REFERENCES

[1] Y.B. Zhang, A. Godfrey, Q. Liu, W. Liu, D. Juul Jensen, "Analysis of the growth of individual grains during recrystallization in pure nickel", **Acta Materialia**, Vol. 57, No. 9, 2009, pp. 2631-2639.

[2] H.K.D.H. Bhadeshia, R. Honeycombe, **Steels: Microstructure and Properties**, Third edition, Elsevier Ltd, 2006.

[3] O.M. Ivasishin, S.V. Shevchenko, N.L. Vasiliev, S.L. Semiatin, **Materials Science and Engineering A**, Vol. 433, 2006, pp. 216-232.

[4] B. Radhakrichnan, G.B. Sarma, T. Zacharia, "Modeling the kinetics and microstructural evolution during static recrystallization—Monte Carlo simulation of recrystallization", **Acta Materialia**, Vol. 46, 1998, pp. 4415-4433.

[5] D. Fan, L.-Q. Chen, "Computer simulation of grain growth using a continuum field model", **Acta Materialia**, Vol. 45, 1997, pp. 611-622.

[6] M. Wang, B.Y. Zong, G. Wang, "Grain growth in AZ31 Mg alloy during recrystallization at different temperatures by phase field simulation ", **Computational Materials Science**, Vol. 45 , 2009, pp. 217-222.

[7] K. Binder, "Time-dependent Ginzburg-Landau theory of nonequilibrium relaxation", **Phys. Rev. B**, Vol. 8, 1973, pp. 3423-3438.

[8] A. Kazaryan, Y. Wang, S.A. Dregia, B.R. Patton, "Grain growth in systems with anisotropic boundary mobility: Analytical model and computer simulation", **Phys. Rev. B**, Vol. 63, 2001, no. 184102.

[9] W.T. Read, W. Shockley, "Dislocation models of crystal grain boundaries", **Physical Review**, Vol. 78, 1950, pp. 275-289.

[10] A. Mallick, S. Vedantam, "Phase field study of the effect of grain boundary energy anisotropy on grain growth", **Computational Materials Science**, Vol. 46, 2009, pp. 21-25.

[11] D. Wolf, "A read-shockley model for high-angle grain boundaries", **Scripta Metallurgica**, Vol. 23, 1989, pp. 1713-1718.

[12] M. Asle Zaeem, H. El Kadiri, P.T. Wang, M.F. Horstemeyer, "Investigating the effects of grain boundary energy anisotropy and second-phase particles on grain growth using a phase-field model", **Computational Materials Science**, Vol. 50, 2011, pp. 2488-2492.

[13] COMSOL Multiphysics Users's Guide, **COMSOL**, Inc., 2010.

Comparison of Cellular Automaton and Phase-Field Models to Simulate Dendritic Solidification

Mohsen Asle Zaeem

Center for Advanced Vehicular Systems, Mississippi State University

Starkville, MS 39759, USA

* mohsen@cavs.msstate.edu, +1 662-325-0126

Hebi Yin

Oak Ridge National Laboratory

Oak Ridge, TN 37831, USA

Sergio D. Felicelli

Mechanical Engineering Department and Center for Advanced Vehicular Systems

Mississippi State University, Starkville, MS 39759, USA

ABSTRACT

In this work, a cellular automaton (CA)–finite element (FE) model and a phase-field (PF)–FE model were developed to simulate dendritic solidification of both cubic and hexagonal crystal materials. Validation of the both models was performed by comparing the simulation results to the analytical model developed by Lipton-Glicksman-Kurz (LGK), showing qualitatively good agreement in the tip growth velocity at a given melt undercooling. Dendritic solidification in cubic materials is illustrated by simulating the solidification in aluminum alloy Al-3wt%Cu. Results show that both models successfully simulate multiple arbitrarily-oriented dendrites for cubic materials. Application to magnesium alloy AZ91 (approximated with the binary Mg-8.9wt%Al), illustrates the difficulty of modeling dendrite growth in hexagonal systems using CA–FE regarding mesh-induced anisotropy and a better performance of PF–FE in modeling multiple arbitrarily-oriented dendrites.

Keywords: Cellular automaton, Phase-field model, Finite element, Dendrite growth, Aluminum alloy, Magnesium alloy.

1. INTRODUCTION

Dendritic solidification process forms the microstructures and has significant effects on the material and mechanical properties of cast alloys [1]. Multi scale nature of solidification patterns which depend on temperature distribution, solute concentration, capillary forces, and kinetic length [2], makes the understanding and predicting these patterns exceptionally difficult. In addition to the experimental solidification research, recently numerical simulations have emerged as a powerful tool in predicting microstructural evolution during solidification. Different mathematical approaches have been employed to investigate the dendritic growth during solidification such as phase field (PF) [3-6] and cellular automaton (CA) [7-9] methods.

In this work, we developed and compared a cellular automaton (CA)–finite element (FE) model and a phase-field (PF)–FE model to simulate equiaxed dendritic solidification in Al-3wt.% Cu and Mg-8.9wt.% Al alloys.

2. CELLULAR AUTOMATON MODEL

The 2D transient equation governing the heat transfer with liquid-solid phase change is given by:

$$\frac{\partial T}{\partial t} = \alpha \cdot \nabla^2 T + \frac{L}{C_\mathrm{p}} \frac{\partial f_\mathrm{s}}{\partial t} \ , \qquad (1)$$

where T is temperature, t is time, α is thermal diffusivity, L is latent heat of solidification, C_p is specific heat, and f_s is the volume fraction of solid. A constant temperature gradient boundary condition is imposed on the four walls. Each finite element is refined into m×m cells for the solute and CA calculation as shown in Fig. 1(b). The temperature of each cell is obtained by interpolating the temperatures of the four nodes of the finite element. The equation governing solute transport is:

$$\frac{\partial C_i}{\partial t} = D_i \cdot \nabla^2 C_i + C_i \cdot (1-k)\frac{\partial f_s}{\partial t}, \quad (2)$$

where C is solute concentration (wt. %), D is the solute diffusivity, k is the partition coefficient and the subscript i indicates solid or liquid. k is the partition coefficient. At the solid/liquid interface, the solute partition between liquid and solid is calculated by

$$C_S^* = k \cdot C_l^*, \quad (3)$$

where C_S^* and C_l^* are the interface solute concentrations in solid and liquid phases, respectively. Eq. (1) and Eq. (2) were solved by finite element method using quadrilateral bilinear elements for Al-3wt.% Cu and hexagonal elements for Mg-8.9wt.% Al alloy.

Kinetics Parameters for CA Model

Based on the local actual liquid concentration C_l calculated from Eq. (2) and the interface equilibrium composition C_l^*, the increase of solid fraction (Δf_s) at interface cells can be obtained.

$$\Delta f_\mathrm{s} = \left(C_\mathrm{l}^* - C_\mathrm{l}\right) / \left(C_\mathrm{l}^* \cdot (1-k)\right), \quad (4)$$

The interface equilibrium composition is calculated by:

$$C_\mathrm{l}^* = C_0 + \frac{T^* - T_\mathrm{l}^{\mathrm{eq}} + \Gamma K \sigma(\theta, \theta_0)}{m_\mathrm{l}}, \quad (5)$$

C_0 is initial solute concentration, $T_\mathrm{l}^{\mathrm{eq}}$ is the equilibrium liquidus temperature at initial solute concentration, T^* is the interface equilibrium temperature, m_l is the liquidus slope, Γ is the Gibbs-Thomson coefficient and K is the curvature of S/L interface. $\sigma(\theta, \theta_0)$ is a function accounting for the anisotropy of the surface tension, in which θ is the growth angle and θ_0 is the preferential orientation with respect to the horizontal coordinate. A detailed description of the CA-FE model is presented in Ref. [10,11], including treatment of nucleation and capturing of interface cells

3. PHASE–FIELD MODEL

A PF variable, $\phi(\mathbf{x},t)$ is considered to distinguish liquid and solid phases similar to the Kobayashi's model [3]. $\phi(\mathbf{x},t) = 0$ represents the liquid phase and $\phi(\mathbf{x},t) = 1$ represents the solid phase. The Gibbs-Thomson equation for an isotropic surface energy and for a binary alloy is [6]:

$$\frac{1}{\mu|\nabla\phi|}\frac{\partial \phi}{\partial t} = T_\mathrm{m} - T + m_\mathrm{l} C_\mathrm{l} - \Gamma \nabla \cdot \mathbf{n}. \quad (6)$$

where μ is a linear kinetic coefficient, T is the temperature, T_m is the equilibrium melting temperature of the pure substance, m_l is the slope of the liquids line from an equilibrium phase diagram, C_l is the equilibrium concentration in the liquid, Γ is the Gibbs-Thomson coefficient, and $\mathbf{n}$ is the unit normal vector exterior to the solid, $\mathbf{n} = -\nabla\phi / |\nabla\phi|$. A double-well potential is considered for the Gibbs free energy [6]:

$$\phi = \frac{1}{2}\left(1 - \tanh\frac{n}{2\omega}\right), \quad (7)$$

where n is the coordinate normal to the interface and 6ω is the interface thickness where ϕ varies between 0.05 to 0.95. The evolution equation of ϕ, considering the anisotropy introduced in by Kobayashi [3], is:

$$\begin{aligned}\frac{\partial \phi}{\partial t} = &-\mu\Gamma\frac{\partial}{\partial x}\left(\sigma\sigma'\frac{\partial\phi}{\partial y}\right) + \frac{\partial}{\partial y}\left(\sigma\sigma'\frac{\partial\phi}{\partial x}\right) \\ &+ \mu\Gamma\nabla\cdot(\sigma^2\nabla\phi) - \frac{\phi(1-\phi)(1-2\phi)}{\omega^2} \\ &+ \mu(T_\mathrm{m} - T + m_\mathrm{l} C_\mathrm{l})\frac{\phi(1-\phi)}{\omega}, \end{aligned} \quad (8)$$

where:

$$\sigma(\theta,\theta_0) = 1 - \delta\cos[j(\theta-\theta_0)]; \quad \theta = \tan^{-1}\left(\frac{\partial\phi/\partial y}{\partial\phi/\partial x}\right). \quad (9)$$

j is 4 for cubic and 6 for HCP crystals.

Another PF variable, $C(\mathbf{x},t)$, is used to represents the concentration of elements in the binary system. The evolution equation for the concentration is [6]:

$$\frac{\partial C}{\partial t} = \nabla\cdot\tilde{D}\left[\nabla C + \frac{(1-k)C}{1-\phi+k\phi}\nabla\phi\right], \quad (10)$$

where

$$\tilde{D} = D_s + (D_l - D_s)\frac{1-\phi}{1-\phi+k\phi}. \quad (11)$$

The evolution of the temperature field is similar to Eq. (1):

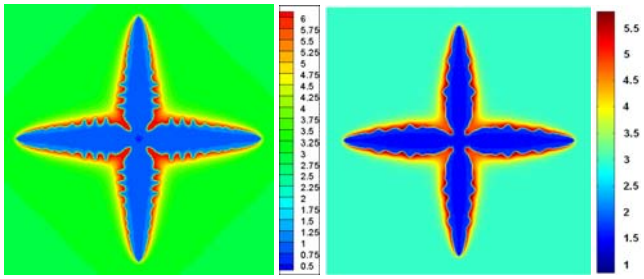
$$\frac{\partial T}{\partial t} = \alpha\cdot\nabla^2 T + \frac{L}{C_\mathrm{p}}\frac{\partial\phi}{\partial t}. \quad (12)$$

A numerical algorithm similar to the previously developed mixed order finite element model [12,13] for solving coupled partial differential equations was adopted in this work. A uniform mesh of square elements with linear interpolation functions was used to discretize the domain for Eq. (8), Eq. (10) and Eq. (12). COMSOL Multiphysics software is used for the computations [14].

3. RESULTS

The dendrite growth of a binary Al-3wt. % Cu alloy and Mg-8.9wt.% Al alloy (approximation of MgAZ91 alloy) was simulated in this study. The materials properties of these allys used in this model can be found at [10,11,15].

In Fig. 1, CA–FE and PF–FE simulation results are plotted for the solidification of a single seed with preferred growth orientations of 0 degree with respect to the horizontal direction. This figure shows a perfect 4-fold symmetry for the primary arms expected for Al-3wt. % Cu alloy.



(a) CA–FE model (b) PF–FE model.

Fig. 1. Solute map at 0.01s simulation time for a single seed with preferred growth orientations of 0 degree with respect to the horizontal direction and with an applied temperature gradient of 400 K/m imposed at the boundaries. (Solidification of Al-3wt. % Cu alloy).

To validate our simulations, we compared our simulation results to the analytical model developed by Lipton-Glicksman-Kurz (LGK) and to the numerical model by Zhu [15]. Fig. 2 shows that results of both CA–FE and PF–FE models are in good agreement with both LGK and Zhu models in the tip growth velocity at a given melt undercooling.

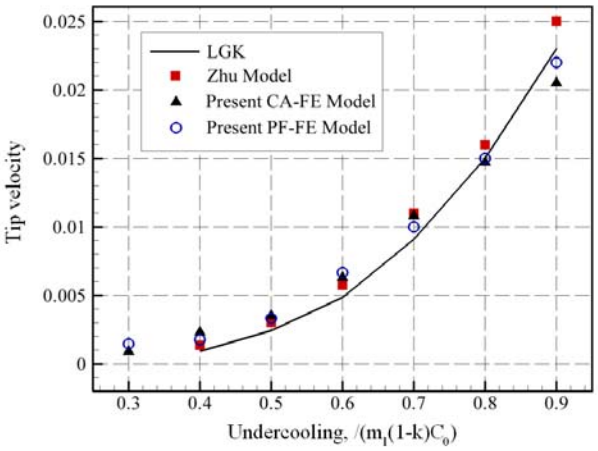
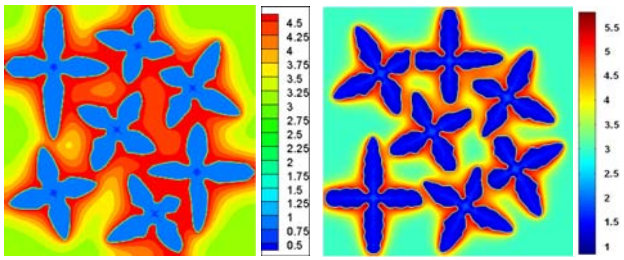


Fig. 2. Tip growth velocity versus undercooling: comparison between numerical simulations and LGK model predictions of the steady-state.

We used a system including 7 arbitrarily-oriented nucleuses to examine the capability of both methods to simulate dendrites growing in random orientation in solidification of Al-3wt. % Cu alloy. Simulation results for CA–FE and PF–FE models are presented in Fig. 3, which show both models could simulate dendrite growth from arbitrarily-oriented nucleuses in cubic crystals.

In the case of solidification of Mg-8.9wt.% Al alloy, both CA–FE and PF–FE models could perfectly simulate the dendrite growth with preferred growth orientations of 0 degree, but CA–FE was not capable of modeling multiple arbitrary-oriented dendrites with six-fold symmetry because of the mesh induced anisotropy resulted from the hexagonal finite elements. The PF–FE simulation results of multiple arbitrary-oriented dendrites with six-fold symmetry are presented in Fig. 4.



(a) CA–FE model (b) PF–FE model

Fig. 3. Solute map of multiple arbitrarily-oriented seeds at 0.014s simulation time (solute transport only). (Solidification of Al-3wt. % Cu alloy).

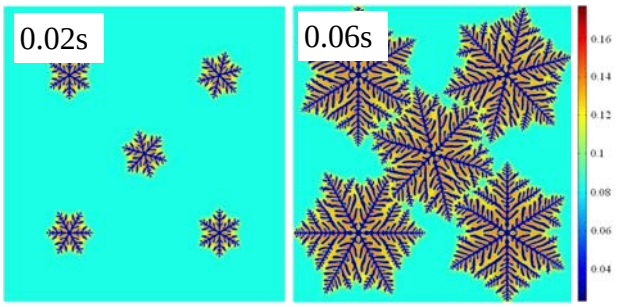


Fig. 4. Solute map of multiple-arbitrary oriented seeds at two simulation times using PF-FE model. Heat flux imposed at the four walls was 10 kWm^{-2}.

4. CONCLUSIONS

A cellular automaton (CA)–finite element (FE) model and a phase-field (PF)–FE model were developed to simulate equiaxed dendritic growth during the solidification of both cubic and hexagonal crystal materials. Models were validated by comparing the simulation results with the LGK analytical model. For cubic crystals with 4-fold symmetries, both models could simulate the dendritic solidification for Al-3wt.% Cu. Also results showed that both models successfully simulate multiple arbitrarily-oriented dendrites for cubic materials. In modeling dendrite growth in hexagonal systems, CA–FE encountered mesh-induced anisotropy

and could not simulate multiple arbitrarily-oriented dendrites. A better performance of PF–FE in modeling multiple arbitrarily-oriented dendrites was concluded.

5. REFERENCES

[1] D. Askeland, **The science of materials and engineering**, Thompson, Brooks/Cole, Belmont, CA, 2003.

[2] R. Trivedi, W. Kurz, **International Materials Review**, Vol. 39, 1999, pp. 49-74.

[3] R. Kobayashi, **Physica D**, Vol. 63, 1993, pp. 410-423.

[4] J. A. Warren, W. J. Boettinger, C. Beckermann, A. Karma, **Annual Review Materials Research**, Vol. 32, 2002, pp. 163-194.

[5] A. Karma, W.-J. Rappel, **Physical Review E**, Vol. 57, 1998, pp. 4323-4349.

[6] C. Beckermann, H.-J. Diepers, I. Steinbach, A. Karma, X. Tong, **Journal of Computational Physics**, Vol. 154, 1999, pp. 468-496.

[7] M. Rappaz, Ch. A. Gandin, **Acta Metallurgica et Materialia**, Vol. 41, 1993, pp. 345-360.

[8] L. B. Sanchez, D. M. Stefanescu, **Metallurgical and Materials Transactions A**, Vol. 35, 2004, pp. 2471-2485.

[9] M. F. Zhu, D. M. Stefanescu, **Acta Materialia**, Vol. 55, 2007, pp. 1741-1755.

[10] H. Yin, S.D. Felicelli, L. Wang, **Acta Materialia**, Vol. 59, 2011, pp. 3124–3136

[11] H. Yin, S.D. Felicelli, **Modelling and Simulation in Materials Scicne and Engineering**, Vol. 17, 2009, No. 075011, pp. 1-15.

[12] M. Asle Zaeem, S. Dj. Mesarovic, **Journal of Computational Physics**, Vol. 229, 2010, pp. 9135-9149.

[13] M. Asle Zaeem, S. Dj. Mesarovic, **Solid State Phenomena**, Vol. 150, 2009, pp. 29-41.

[14] COMSOL Multiphysics Users's Guide, **COMSOL**, Inc., 2010.

[15] D. Sun D, M. Zhu, S. Pan, D. Raabe, **Acta Materialia**, Vol. 57, 2009, pp. 1755-1767.

Advanced Corrosion Resistance applied to Space Operations

S. Curatolo CZT Inc. Lawrence KS USA
Susana.curatolo@cztinc.com

Abstract

It is well known that materials subject to corrosive agents undergo a deteriorating process which includes space operations. For this reason the reliability and reusability of materials such as metals and polymers in space need to be corrosion-resistant. In this paper an attempt is made to show that not only the surfaces but the whole bulk can be made to undergo changes which involve removal of all type of defects creating long-range order within the material structure. This process in turn increases the melting point by 1/3 and thereby rendering the material resistant by 1/3 to amorphisation, as well as the near-diamond hardness with creep resistance, fracture toughness and oxidation resistance. These results have been previously reported as being due to a specific treatment, which has been termed OHRS (Oxygen Hydrogen Resistant Silicon). The suggestion is made that this treatment of materials be directly applicable to the different stages inspace operations. This includes: on earth, traversing the atmosphere including LEO (low earth orbit)and space as such. During the traversing process some major difficulties encountered are cracking due to vibration (vortex shedding) enhanced by defects in the material, ATOX (atomic oxygen attack) as well as micrometeoroids during LEO, charged particle radiation up to the Clarke orbit and beyond.

Introduction

In previous work[1-5] the fundamental nature of matter was investigated as expressed in materials like metals and polymers. Specifically, the research done involved silica as applied to fiber optics and steel alloys as applied to the oil industry. The basic results encompassed the fundamental properties of the individual materials. In the case of fiber-optics it involved the no loss transmission and absorption of light in high-speed silica fiber as related to the defects in the surface and bulk of the silica material. In the case of steel alloy it involved the resistance to corrosion by napthenic acid as related to the defects in the surface and bulk of the steel material. Fundamentally the method deals with the removal of defects based on preventing corrosion by such constituents as oxygen, hydrogen and hydroxyl in silicates and steel alloys. Because corrosion is a serious problem in all stages of space operations, and because of the complex nature of detecting corrosion in the different types of materials and structures constituting a space operation, and because of the natural votive of space flight, the fundamental underpinning root of the corrosion problem is in the material. It poses a fundamental limit.

Experimental Data

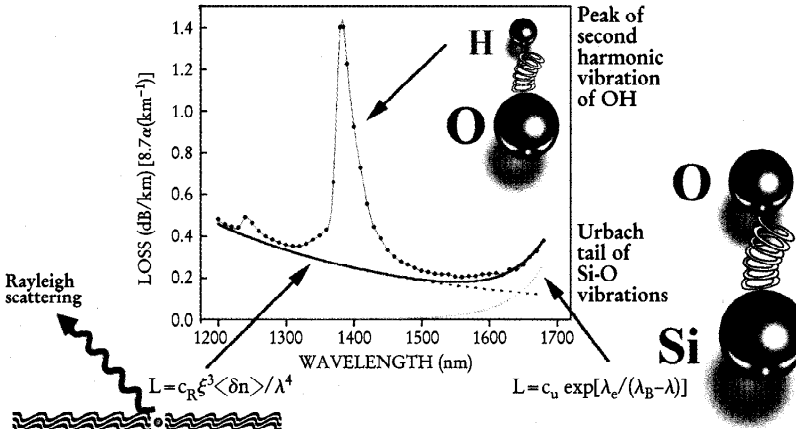
I. OHRS versus the fundamental limit in Silicates

OHRS is a hydroxyl and hydrogen resistant silicon based material. The process uses OHRS and teardrop-geometry hohlraum to render any bulk material OH and H corrosion resistant. Because of this, no hydroxyl or

hydrogen overtone vibrations occur and therefore there are no absorption losses. This process eliminates all defects from the surface and the bulk : insterstitial defects, color defects and pores. and therefore, no defect vibrations occur, no absorption and no dispersion. No dangling bonds. The absorption of the OH, H and water causes the breakdown of the silicate and the metal.
Since all materials have a certain degree of disorder (defects) and additional defects are introduced from the processing of the silicate and the metal, and these openings in the lattice of the structure is where the OH and H bind as a polar bond or interaction or elastic collision.

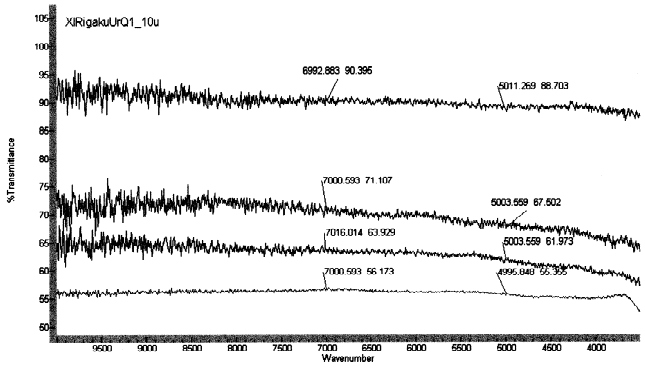
I. Material Structural Changes

Transmission loss in silica has three key components: OH absorption (producing two peaks, one big peak at approximately half the wavelength λ of the fundamental OH mode (λo) and a small peak at slightly shorter λ wavelength; Rayleigh scattering proportional to the inverse of the 4th power of wavelength; and the Urbach exponential tail of Si-O vibrations.
The Clarity Limit of SiO2 material is the sum of the Rayleigh and Urbach contributions. See Figure 1. CZT OHRS (Oxygen,Hydroxyl,Hydrogen,H20 resistant silicon), shows that its raw and un-normalized transmission and absorption performance is flat between 1000nm and 3000+nm, with no absorption peaks and a refraction index difference equal to the uncertainty of the results shown, which amounts to 80% full transmission. See Figure (2). The optical loss profile is 0.003969 dB/Km, and it takes into account the additive term of the log of the length traveled by the light divided by the length. The corresponding power loss is 21%, with power out as a % of power in of 79%. OHRS is a long range order and Input/Output Intensity wavelength independent material. OHRS has a unique structure, and can be pressed, melted or pulverized, keeping its unique micro crystalline aggregate configuration. It is a high k material. The clarity of the transmission medium is the most important factor in determining the practicality of optical communications. Crystal clear transmission without loss due to OH vibration, H vibration and Si-O vibration is possible due to CZT's OHRS material and process. See Figure 1. Additionally, the process renders the solid near-diamond hard (>95Rockwell), physically and chemically corrosion resistant with 2000°+Celcius tolerance.



Figure(1) Absorption Loss in Silica

Figure(2) Flatness-Detail 1000-3000nm



II. OHRS vs The Fundamental Limit in steel

In order to address the needs of the oil industry, especially regarding low carbon alloy steel corrosion and environmental protection, tests were performed on two types of industry steel which include J55C and A516-70. Under very high acidic attack test conditions known in the industry as Napthenic Acid Corrosion at TAN 4.3 , test results showed that OHRS treated steel samples did not corrode as shown in Fig(3). The very low sulfur content of the oil, exacerbated the acid attack at static conditons for one week, and at16ft/s flow condition for another week. The centrifuge temperature was at fractioning level of 600˚F, and 600psi back pressure. One population of samples were subjected to severe thermal shock at 750°C prior to acidic oil test. Therefore, the OHRS material and treatment method is a natural solution to preventing corrosion damage, and an excellent, environmentally safe, and non-expensive alternative to other treatment methods, noxious inhibitors or cathodic and electrochemical methods. Considering that the typical NACI for carbon steel is 22.5, and that its mpy is 15-18 [13], whereas OHRS treated steel shows 0 mpy, See Fig(3)-Sample 4 Fig4(a,b) for SEM results show the J55C and A56-70 with no defects on the surface.

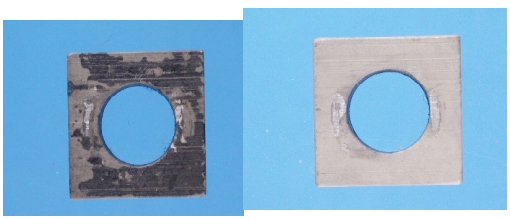
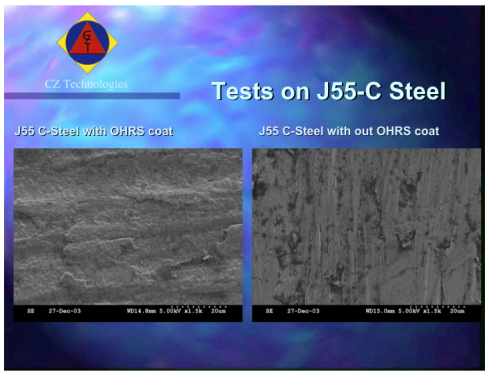
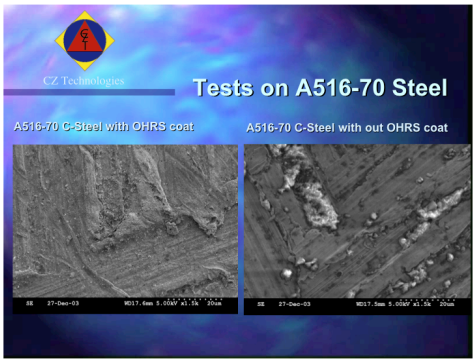


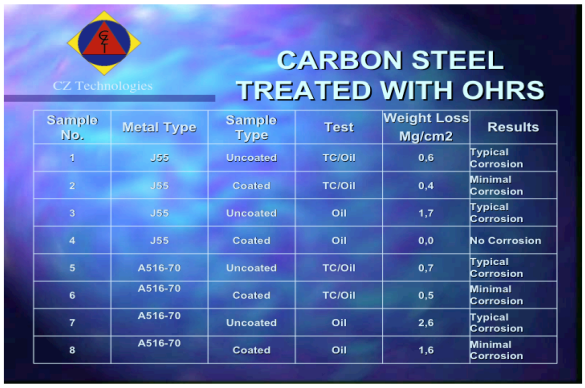
Figure (3) Left-untreated,Right-Treated





Fig(4)(a,b)SEMResolution,(a)(Left-Treated;Right-untreated);(b)Left-treated;Right-untreated) Res: $1\mu m$

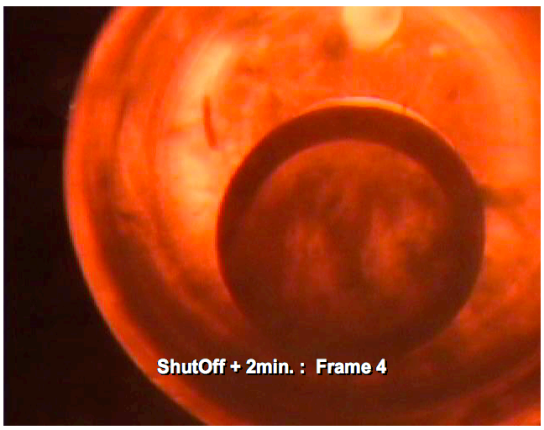
Table(1) Corrosion rates for 2 CZT treated/non-treated low alloy steels (1.25Cr)atTAN=4.3/300°C/low-S/16ft/s/ 600psi/Pre-test thermal Shock 30 min 750°C: Sample 4 = 0 mpy [14]



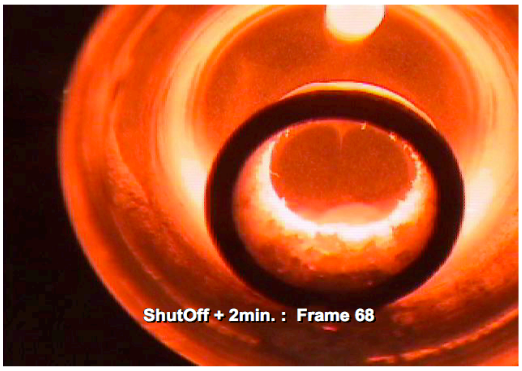
Sample No.	Metal Type	Sample Type	Test	Weight Loss Mg/cm2	Results
1	J55	Uncoated	TC/Oil	0,6	Typical Corrosion
2	J55	Coated	TC/Oil	0,4	Minimal Corrosion
3	J55	Uncoated	Oil	1,7	Typical Corrosion
4	J55	Coated	Oil	0,0	No Corrosion
5	A516-70	Uncoated	TC/Oil	0,7	Typical Corrosion
6	A516-70	Coated	TC/Oil	0,5	Minimal Corrosion
7	A516-70	Uncoated	Oil	2,6	Typical Corrosion
8	A516-70	Coated	Oil	1,6	Minimal Corrosion

III. The Effect of the teardrop-geometry hohlraum

A test was performed on the effectiveness of the teardrop-geometry hohlraum by using elemental gold as the agent under the following experimental conditions. The hohlraum is substantially-spherical-teardrop geometry, and can be made of fused quartz, metal or other materials.The solid is placed in the envelope and evacuated. The envelope is placed in an oven and heated. This experiment is proof that these conditions will produce a self generated magnetic field which is at the basis of all energetic processes within the Hohlraum. Under normal circumstances, when the heat energy is added to the material, electrons (p-electrons,[2,4] Curatolo EPS London June 2004) are excited and leave the surface obeying Einstein's photoelectric effect. The electrons (negatively charged particles) will leave the close proximity of the material and the material left behind becomes positively charged. The motion of the electrons emitted from the solid surface are trying to come into equilibrium following the shape of the Hohlraum (Curatolo EPS London June 2004, Patent #2)[2,4]. The total motion of the electrons creates a magnetic field parallel to the axis of motion, and due to the Lorentz force on each emitted electron moving around in the fixed volume, additional energy is generated and added to that of the heat flow. The result is shown in Figure 5 (a-b) which were filmed by video camera. A Dipole Field is clearly visible.



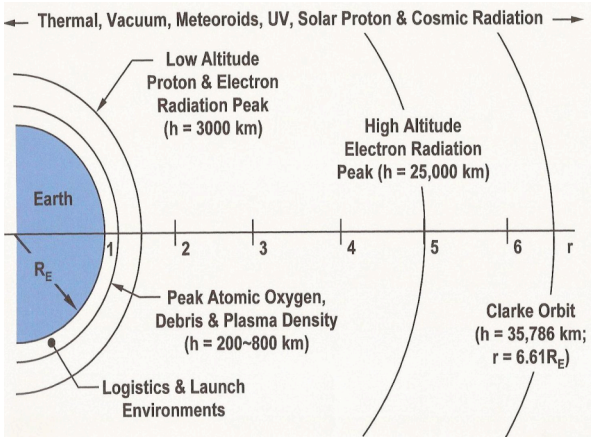
Figures (5) a



Figures (5) b

Application and Discussion

Based on the data presented above we attempt in this paper to indicate the significance of this data applied to the corrosion condition encountered in space flight. First let us recall the most likely conditions encountered in space operations. We may consider mainly three stages, on the ground, LEO and the Clarke orbit.



Figure(6) Spatial Distribution of Environments at the equator from earth surface to Clarke orbit with Overall converted to mpy erosion rates for Metal/polymer [6] (Graph Adapted from ETA Flight materials Wakefield MA, Ray Erikson, 8 Nov 2001, ray@eriksontech.com)

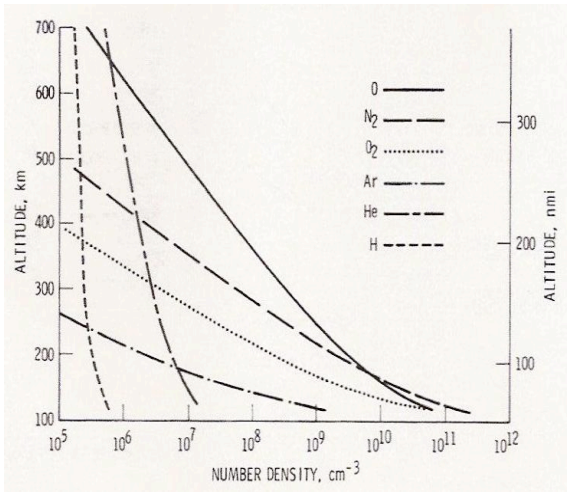


Figure (7) Atmospheric composition in low earth orbit (LEO)[7,8] (From: Corrosion in Space,Rooij,A.ESA,Noorwijk,The Nether-lands).

Table (2) Corrosion comparison in mpy

Grade metal/Polymer	Corrosion rate (mpy)
OHRS-metal (Ground)	0
OHRS-polymer (Ground)	0
Commercial-metal (Ground)	18
Earth-Commercial-polymer [12]	5
Space-Commercial metal/polymer(LEO) [6,9,10]	126
Space-Commercial-Metal-polymer [11] (Ground)	42

From this comparison in Table (1) and Figure (6) and Figure (7), there is clearly an advantage of treating metals and polymers as suggested by the CZT OHRS method before ant deployment to space is considered.

Conclusion

In view of the fact that the OHRS process is relatively cost effective, the performance advantage to space operation materials, if treated by this method, would be obvious. It should be pointed out, that not only the surface is affected by OHRS, but the complete bulk material as well, making it resistant to the eroding agents encountered during space flight. It seems, that the OHRS bulk treatment would be superior to some of the latest methods of tolerance enhancement, such as nanostructured super alloys (Nb, and Mo) and low angle grain boundary resistant crystal super alloy used in the shuttle and space operations. This is, particularly because they do not remove the defects in the structure once it is subjected to temperatures near or above 800° C and Oxygen and Hydrogen attack.

Aknowledgements: I would like to thank P.A. Curreri for advice and review of the paper, and G.Dreschhoff for discussions

References

[1] Curatolo ,S. USPTO **patent**: 7067006 (Issue date 27 Jun 2006 2004 PCT WO 2004/005203

[2] Curatolo, S.USFiling date: 23 Nov 2005. Issued **patent**: 7214903 (Issue date 8 May 2007

[3]Curatolo, S. Melting and Vaporizing-apparatus:Implications for ICF-American Physical Society, April Meeting, 2004, May 1-4, 2004, Denver, Colorado April 2004, MEETING ID: APR04, abstract #R14.005

[4] Curatolo, S. New Experimental implications for ICF, http://epsppd.epfl.ch/London/pdf/P2_002.pdf, 31rst EPS Conf. Plasma Physics,London 28 june-2nd July,2004 ECA Vol.28GP-2.002 (2004)

[5] Curatolo,S. OHRS-OH Silicon resistant Material-American Physical Society, March Meeting 2004, March 22-26, 2004, Palais des Congres de Montreal, Montreal, Quebec, Canada, MEETING ID: MAR04, abstract #U10.011

[6] Erikson,R. Space environments and effects,Nov 8,2001,ETA Flight Materials, Wakefield MA USA, ray@eriksontech.com

[7] De Rooij, A., Corrosion in Space, Head of Materials technology,Section Materials & Components Technology Div. ESA European Space technology and research center, Noordwijk, The Netherlands.pp1-13

[8] De Rooij, A. ,The degradation of metal surfaces by atomic oxygen, 3rd European symposium of spacecraft materials in Space Environment, ESA,1985, Noordwijk,The Netherlands

[9] Silverman,E.M.;Atomic Erosion yield Measurement, NASA CR-4661

[10] Leger,L.J., Oxygen atom reaction with shuttle materials at orbital altitudes, NASA TM-58246, 1982

[11] Kennedy Space Center-Corrosion technology Laboratory-NASA, ksc-corrosion@mail.nasa.gov

[12]Restek, Bellefonte PA 16823,Cardinal Lab, St. Louis MO 63114; Study of 6N HCl Corrosion on Commercial Silico CR Coatings-316L SS, Hastelloy C-22 & True Tube ™ Variants, April 2004

[13]http://www.ccqta.com/docs/documents/Projects/TAN_Phase_II/TAN%20Phase%20II%20-%20Feb%2006.pdf

[14]CZT Inc. Executive Summary, Crude-oil Opportunity 2004, www.cztinc.com

The Simulation and Analysis of the Closed Die Hot Forging Process by A Computer Simulation Approach

Dipakkumar GOHIL
Department of Mechanical Engineering, S. V. National Institute of Technology
Surat - 395 007, Gujarat state, India

ABSTRACT

The objective of this research work is study the variation of various parameters such as stress, strain, temperature, force, etc. during the closed die hot forging process. A computer simulation modeling approach has been adopted to transform the theoretical aspects in to the computer algorithm which would simulate and analyze the closed die hot forging process. For the purpose of process study, the entire process has been divided in to appropriate finite number of steps and the output values have been computed at each deformation step. The results of simulation have been graphically represented and suitable corrective measures are also recommended if the simulation results do not agree with the theoretical values. This computer simulation approach would significantly improve the productivity and reduce the energy consumption of the overall process for the components manufactured by closed die forging process and contribute towards the efforts in reducing the global warming.

Keywords: Closed die hot forging, H/D ratio, Modeling, Simulation.

1. INTRODUCTION

In the highly competitive present era, the mass production requirements in the engineering industries have increased the demand for the forged components. Forging process, usually involve multiple preforming processes followed by a specified finishing process.

Process simulation has become an increasingly important tool for the development of new or improved processes. The effective use of simulation tools will not only reduced the costs and the time necessary for the development of new products significantly but also helps and contributes in efforts of reducing the global warming effects. The process requires a lot of experience and skill to optimize the quality, costs and lead time.

The objective of this research work is to simulate and analyze the closed die forging process. The main focus is to know the variation of stress, strain, force, temperature, etc. at different stages of forging of AISI 1016 by a computer simulation method. In this approach, the closed die forging process has been divided in to two stages i.e. deformation process before flash formation and after flash formation.

The designs based on results of simulation are required to be evaluated to make sure that the material would flow as per the requirement in the die cavity.

2. MODELING OF CLOSED DIE FORGING

The modeling and simulation of closed die forging process has been carried out by using analytical and numerical methods [1], [12]. For the purpose of simulation, a cylindrical shape billet, having appropriate H/D ratio, have been deformed between two die halves as shown in Fig. 1. The deformation process before flash formation and after flash formation has been depicted in Fig. 2 and Fig. 3 respectively.

Estimation of the Flow Stress

Forging is a deformation process which involves several variables interconnected by more or less complex functions. The method discussed here is used to simulate the hot die forging process to determine the flow stress, maximum stress, strain before flash formation, strain in flash, load, temperatures of die and billet. The flow stress relationship has been implemented as a subroutine [1].

$$\sigma_f = C\dot{\varepsilon}^m \quad (1)$$

Where, $C = f(\varepsilon, T)$

The parameters C and m are available in the material property hand books for different billet materials used in various types of forming processes.

Calculation of Strain

The strains in the die cavity as well as in the flash during different stages have been calculated as follows [4].

The strain before flash formation, as shown in Fig. 2, could be computed as,

$$\varepsilon = \ln(H / h) \quad (2)$$

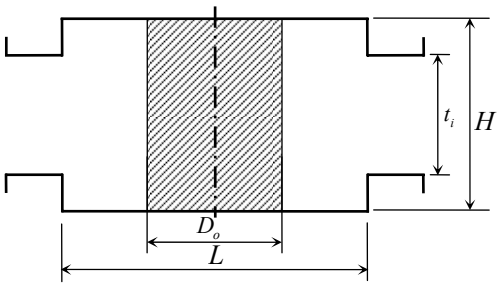


Fig. 1 Billet and Die geometry: Initial configuration

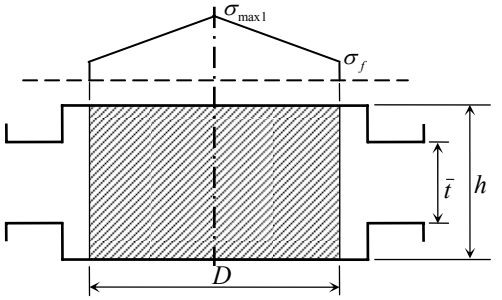


Fig. 2 Billet and Die geometry: Upsetting

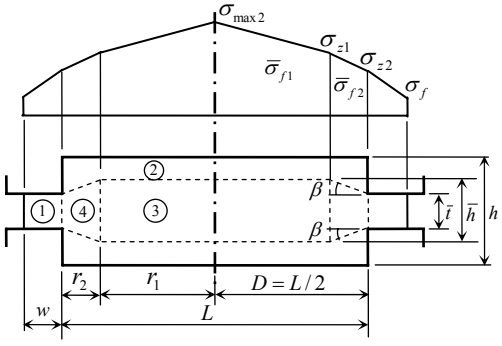


Fig. 3 Billet and Die geometry: with Flash

During flash formation, as depicted in Fig. 3, the strain in flash, indicated as encircled 1 (zone 1), may be calculated as,

$$\varepsilon_1 = \varepsilon_{tc} + \ln(\bar{t}_c / \bar{t}) \tag{3}$$

At the end of the process, the material is assumed to flow in to flash by shearing along the surface indicated by dashed line in Fig. 3. Therefore, the strain in zone 2, shown as encircled digit 2 (zone 2) i.e. ε_2 is equal to ε_{tc} which is the zone between the shearing line and the die, and the strain in zone 3, represented by encircled digit 3, is ε_3 may be computed as,

$$\varepsilon_3 = \varepsilon_{tc} + \ln(\bar{h}_c / \bar{h}) \tag{4}$$

$$\bar{h} = 0.8\bar{t}(L / 2\bar{t})^{0.92} \tag{5}$$

The mean height of the convergence region 4 indicated by encircled digit 4 in Fig. 3, could be calculated as,

$$\bar{h}_1 = (\bar{h} + \bar{t}) / 2 \tag{6}$$

The strain in the convergence region 4 may be obtained as,

$$\varepsilon_4 = \varepsilon_{tc} + \ln(\bar{h}_c / \bar{h}_1) \tag{7}$$

Calculation of Strain Rates

The strain rates for the various stages of closed die forging may be calculated as follows:

Deformation process before flash formation as shown in Fig. 2 may be computed as,

$$\dot{\varepsilon} = v / h \tag{8}$$

Deformation process after Flash formation as depicted in Fig. 3 may be evaluated as,

$$\dot{\varepsilon}_1 = v / \bar{t} \tag{9}$$

$$\dot{\varepsilon}_3 = v / \bar{h} \tag{10}$$

$$\dot{\varepsilon}_4 = v / \bar{h}_1 \tag{11}$$

Calculation of Stresses

The stress before flash formation as shown in Fig. 2 may be computed as,

$$\sigma_{max} = \sigma_f (1 + (\mu_1 D / h)) \tag{12}$$

The stresses after flash formation as represented in Fig. 3 could be calculated as,

$$\sigma_{z2} = 2\mu_1 \sigma_f (w / \bar{t}) + \sigma_f) \tag{13}$$

$$\sigma_{z1} = (K_2 / K_1) \ln(\bar{t} / K_3 + K_2 r_2)) + \sigma_{z2} \tag{14}$$

$$\sigma_{max} = ((2\mu\sigma_{f1} r_1) / h) + \sigma_{z1} \tag{15}$$

Where,

$$K_1 = -2\tan\beta \tag{16}$$

$$K_2 = -\sigma_{f2} K_1 + 2\mu_2 \sigma_{f2} (1 + \tan^2 \beta) \tag{17}$$

$$K_3 = \bar{h} - r_2 K_1 \tag{18}$$

$$\tan\beta = [1 - \{((\bar{h} / \bar{t}) - 1) / ((\bar{h} / \bar{t}) \ln(\bar{h} / \bar{t}))\}]^{1/2} \tag{19}$$

Calculation of Forces

The force required for deformation process before flash formation for the geometry as shown in Fig. 2 could be,

$$F = 2\pi\sigma_f [(h^2 / 4\mu_1^2)(\exp(D / h) - 1) - hD / 4\mu_1)] \tag{20}$$

The force required for deformation process after flash formation for the geometry as shown in Fig. 3 would be,

$$F = F_1 + F_2 \tag{21}$$

Where,

$$F_1 = 2\pi\sigma_{z1} [(\bar{h}^2 / 4\mu_2^2)(\exp(\mu_2 L / \bar{h}) - 1) - \bar{h}L / 4\mu_2)] \tag{22}$$

$$F_2 = \sigma_f (\pi / 4)(2Lw + w^2)(1 + (\mu_1 w / 3\bar{t})) \tag{23}$$

F_1 is the average force on the billet and F_2 is the average force on the flash land.

Calculation of Die and Billet Temperatures

The temperatures during the process are calculated separately within the die cavity and in the flash [2], [14].

Die cavity

The analysis for the temperature distribution for a system comprising of die-lubricant-billet may be summarized in non-dimensional form as follows.

The time has been transformed in to a non-dimensional parameter N_θ as follows.

$$N_\theta = \alpha\theta / H^2 \quad (24)$$

$$\alpha = \lambda / \rho C_P \quad (25)$$

$$FS_{Die} = (T_D - T_{DO})/(T_{BO} - T_{DO}) \quad (26)$$

$$FS_{Billet} = (T_{BO} - T_B)/(T_{BO} - T_{DO}) \quad (27)$$

The FS_{Die} and FS_{Billet} are the two non-dimensional functions representing the change in temperature at different locations within the die and the billet.

For different H/D ratios, the non-dimensional parameter N_θ is related with FS_{Die} and FS_{Billet} as shown in Fig. 4 and Fig. 5 respectively.

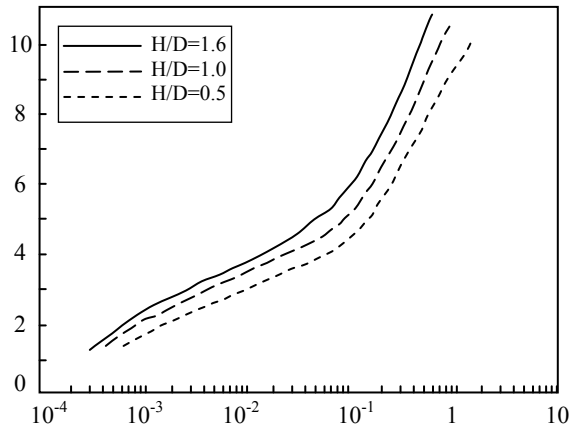


Fig. 4 FS_{Billet} Vs. N_θ

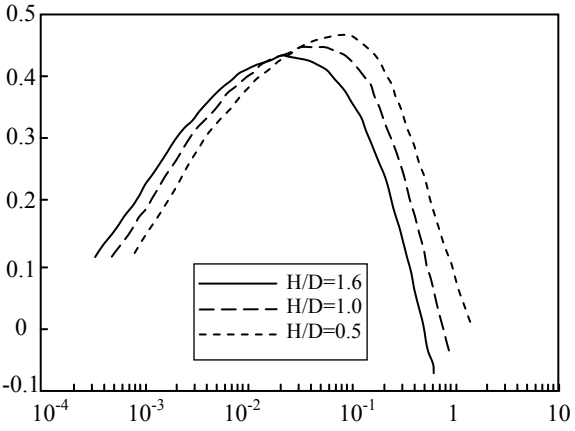


Fig. 5 FS_{Die} Vs. N_θ

The change in temperature with time could be calculated as follows.

$$T_D = T_{DO} + FS_{Die}(T_{BO} - T_{DO}) \quad (28)$$

$$T_B = T_{BO} - FS_{Billet}(T_{BO} - T_{DO}) \quad (29)$$

The temperature rise due deformation may be computed as,

$$\Delta T = \int (\sigma_f / \rho\, C_p)\, d\varepsilon \quad (30)$$

The heat developed during deformation is partially transferred to the die which may be calculated as,

$$T_{Bd} = \Delta T \exp(-\alpha_T\, t / \rho\, C_p\, h) \quad (31)$$

The final temperature of the billet may be given by,

$$T_{Bf} = T_B + T_{BD} \quad (32)$$

Flash land

Due to the relatively small thickness of the flash as compared to the die cavity the temperature gradient may be neglected. Thus, the average flash temperature may be calculated as,

$$T = T_{DC} + (T_{BC} - T_{DC}) \exp(-\alpha_T (t - t_{cont}) / \rho\, C_p\, \bar{t}) \quad (33)$$

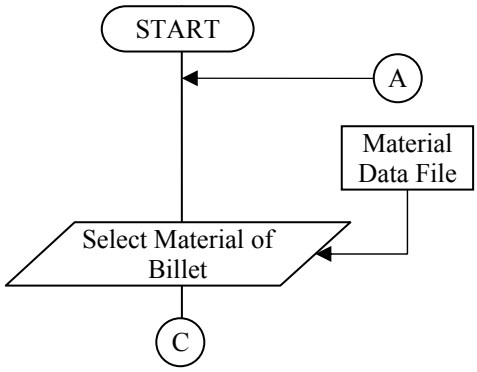
The final temperature in the flash is therefore,

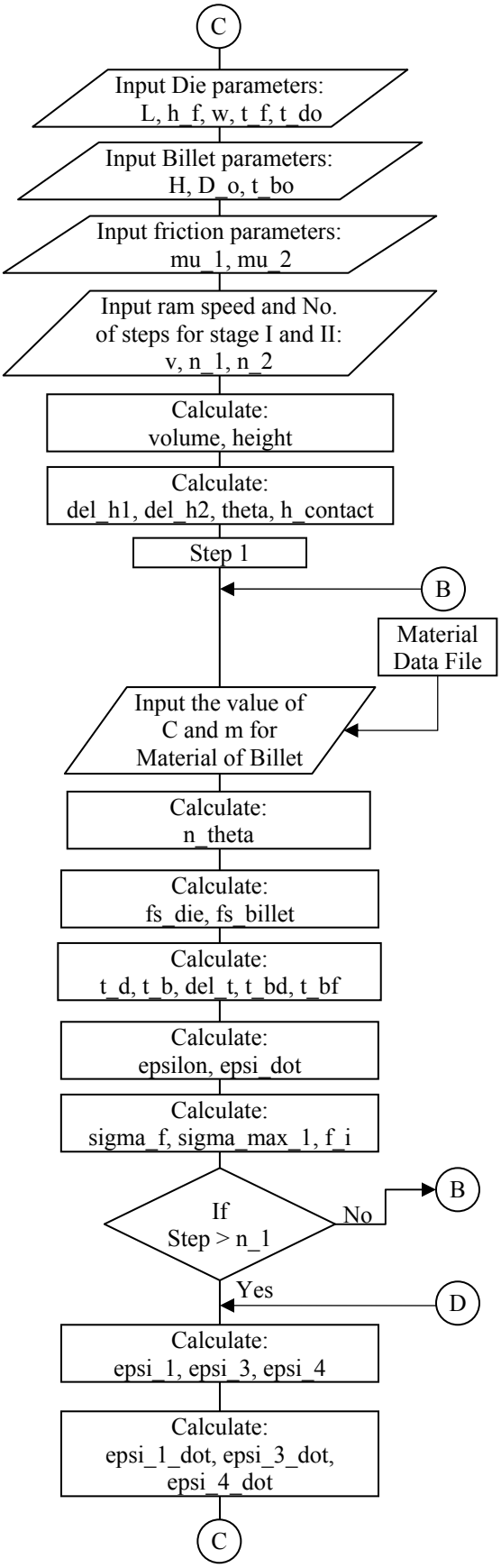
$$T_F = T + \Delta T \quad (34)$$

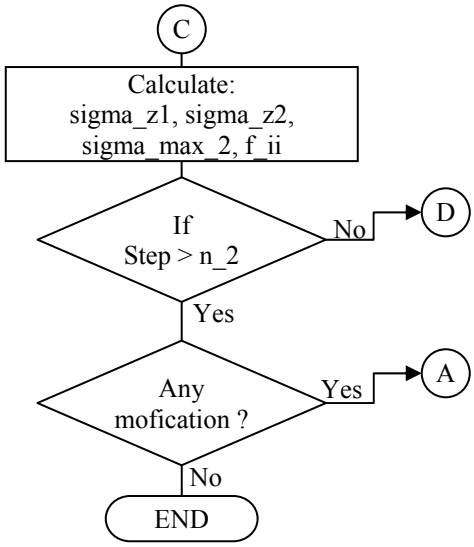
3. COMPUTER ALGORITHM

A computer simulation algorithm has been developed consisting of separate subroutines for the two different stages of forging process i.e. deformation process before flash formation and after flash formation [5], [8].

For the process simulation, initially the material is required to be selected from the material data base. The die parameters, billet parameters, number of stages of deformation process are required to be provided then the algorithm starts computations and simulations using computer graphics as per coding shown below [9], [10].







The computer code generates the results step by step for the various output parameters e.g. height, diameter, strain, strain rate, flow stress, force, temperature of die and billet. At the end of the simulation, the algorithm asks for any modifications, if any, may be change of material or dimensions, for the purpose of next computations and simulation.

The application of this computer simulation method may be extended to more complicated parts such as a connecting rod of an Internal Combustion Engine and aero-engine components [6], [7], [12].

4. RESULTS AND DISCUSSIONS

Fig. 6 shows the graph of flow stress and maximum stress plotted against the deformation steps for the billet H/D ratio of 1.23 and flash thickness of 2 mm. From the Fig. 6, it could be clearly seen that, as the deformation proceeds, the flow stress in the material steadily increases and the value of maximum stress also gradually increases. After a particular step, the increase in the maximum stress is steep; this is the point where the die cavity is almost filled and the flash formation starts.

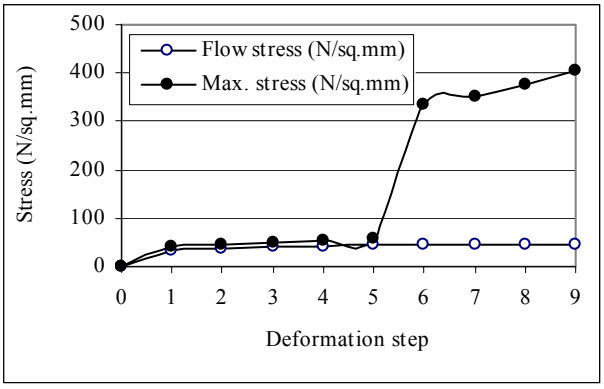


Fig. 6 Flow stress and Max. stress Vs. Deformation

Fig. 7 depicts the graphical representation of strain of billet plotted against the deformation steps. From the Fig. 7, it may be seen that, with the progress of deformation process, the strain in the material increases. During initial process of deformation i.e. deformation before flash formation, the increase in strain is more as compared to the deformation during the flash formation.

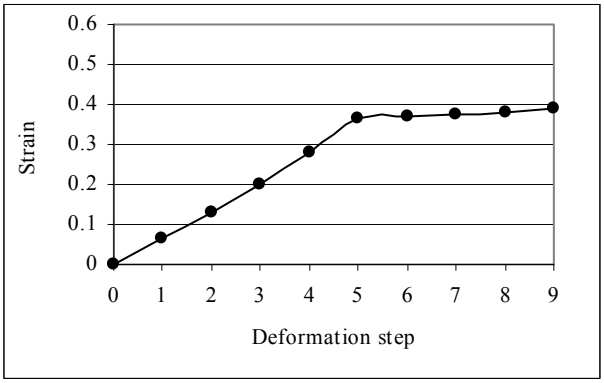


Fig. 7 Strain of billet Vs. Deformation

Fig. 8 shows the graph of variation of strain against the deformation steps during the flash formation process. The flash formation process starts when the die cavity gets almost filled and the extra material comes out resulting in to the flash formation. From the Fig. 8, it could be seen that, with the progress of deformation process, the strain in the flash continuously increases.

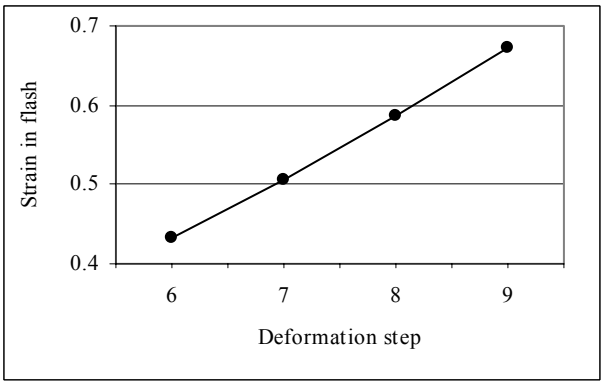


Fig. 8 Strain in flash Vs. Deformation

Fig. 9 represents the graph of variation in the die temperature and billet temperature against the deformation steps. From the Fig. 9, it could be clearly seen that, as the deformation proceeds, the billet temperature continuously keeps on decreasing and on the other side the die temperature keeps on increasing. This change in the temperatures of die and billet occurs in such a way that after some period of time, both, a die and a billet shall achieve the same temperature.

From the Fig. 10, it could be clearly seen that, as the stroke proceeds, the load requirement for the deformation increases. At a particular stroke length, the increase in the load is steep; this is the point where the flash formation starts.

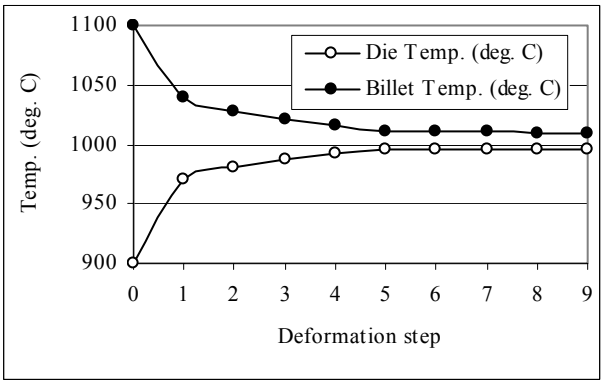


Fig. 9 Die and Billet temp. Vs. Deformation steps

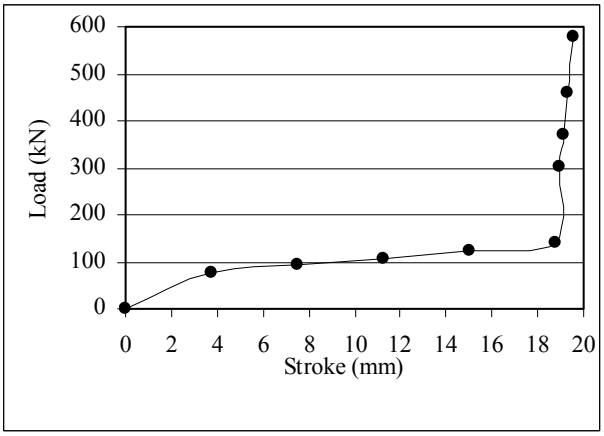


Fig. 10 Load (kN) Vs. Stroke (mm)

5. CONCLUSIONS

In light of above graphical representations, the variation of the parameters such as flow stress, maximum stress, strain before flash formation, strain in flash, force, temperatures of die and billet have been found in good agreement with reference to their theoretical aspects discussed at the process modeling.

Finally, it may be concluded that the success of the simulation method would depend upon the following:

- The selection of material parameters from the material property database should match with the actual material. If the material properties are not known then it may be obtained by the suitable material testing method.
- The friction conditions considered in the computer simulation should be as close as possible to the conditions prevailing between the die and billet material during the actual deformation process.

However, due to the complexity of forging operations, the material and process conditions, the manufacturing by

forging process is still a very much dependent upon trial runs, which results into increased lead-time and cost. An integrated approach of combining the data obtained by the computer simulation method with that of the few trial runs could greatly optimize the closed die hot forging process and also contributes towards the efforts in reducing the global warming.

6. NOMENCLATURE

C	Strain hardening constant
C_P	Specific heat of the material
D_o	Initial diameter of billet
F_I	Force required before flash formation
F_{II}	Force required after flash formation
H	Initial height of billet
$\bar{h}$	Current height of billet
$\bar{h}_c$	Average height of shearing zone at the beginning of flash formation
$\bar{h}_1$	Current average height of shearing zone
L	Die diameter
m	Strain rate sensitivity exponent
t_i	Initial distance between flash lands of two die halves
t	Deformation time
t_{cont}	Time of contact or start of flash formation
$\bar{t}$	Flash thickness
$\bar{t}_c$	Flash Thickness at the beginning of flash formation
T_{BO}	Initial billet temperature
T_B	Current billet temperature
T_{Bd}	Temperature rise of the billet due to deformation
T_{DO}	Initial die temperature
T_D	Current die temperature
ΔT	Increase of temperature due to deformation
v	Ram speed
w	Flash width
α_T	Heat transfer co-efficient between die and billet
β	Shear plane angle
ε	Strain before flash formation
ε_1	Strain for zone 1
ε_3	Strain for zone 3
ε_4	Strain for zone 4
$\dot{\varepsilon}$	Strain rate before flash formation
$\dot{\varepsilon}_1$	Strain rate for zone 1
$\dot{\varepsilon}_3$	Strain rate for zone 3
$\dot{\varepsilon}_4$	Strain rate for zone 4
ε_{tc}	Strain at the beginning of the flash formation
θ	Time in seconds
λ	Thermal conductivity of the billet material
μ_1	Coefficient of friction between die and billet
μ_2	Coefficient of shearing friction
ρ	Density of the material
σ_f	Flow stress
σ_{max1}	Maximum stress before flash formation
σ_{max2}	Maximum stress after flash formation

Encircled digits 1, 2, 3, and 4 indicate respective zones in Fig. 3.

7. REFERENCES

[1] Altan T. and Boulger F. W., "Flow stress of metals and it's application in metal forming analysis", **Trans. ASME**, Series B, Vol. 95, pp. 1009-1018, 1973.

[2] Alexandre Polozine and Lirio Schaeffer, "Influence of the inaccuracy of thermal contact conductance coefficient on the weighted-mean temperature calculated for a forged blank", **Journal of materials processing technology**, pp. 260–266, 195 (2008).

[3] B. Tomov, R. Radev, and V. Gagov, "Influence of flash design upon process parameters of hot die forging", **Journal of Materials Processing Technology**, pp. 620–623, 157–158 (2004).

[4] Biner S. B., "A procedure for determination of the strain distribution during simulation of metal forming using model material techniques", **Journal of Engineering for Industry**, Vol. 114, pp. 94-99, Feb. 1992.

[5] Boër C. R., Rebelo N., Rydstad H., and Schröder G., **Process Modeling of Metal Forming and Thermo-mechanical Treatment**, Springer - Verlag, Berlin, Heidelberg, 1986.

[6] H. Grass, C. Krempaszky, and E. Werner, "3-D FEM-simulation of hot forming processes for the production of a connecting rod", **Computational Materials Science**, pp. 480–489, 36 (2006).

[7] H. Ou, J. Lan, C.G. Armstrong, and M.A. Price, "An FE simulation and optimisation approach for the forging of aeroengine components", **Journal of Materials Processing Technology**, pp. 208–216, 151 (2004).

[8] Hyunkee Kim, Kevin Sweeney, and Taylan Altan, "Application of computer aided simulation to investigate metal flow in selected forging operations", **Journal of Materials Processing Technology**, pp. 127-154, 46 (1994).

[9] M. Bakhshi-jooybari, I. Pillinger, P. P. Hartley, and T. A. Dean, "Finite element simulation and experimental study of hot closed-die upsetting", **International Journal of Machine Tools Manufacturing**, Vol. 36, No. 9, pp. 1021-1032, 1996.

[10] Markus Knoerr, Joon Lee, and Taylan Altan, "Application of the 2D finite element method to simulation of various forming processes", **Journal of Materials Processing Technology**, pp. 31-55, 33 (1992).

[11] Rusinoff S. E., **Forging and Forming Metals**, American Technical Society, Chicago, USA, 1959.

[12] Shinichiro Fujikawa, "Application of CAE for hot-forging of automotive components", **Journal of Materials Processing Technology**, pp. 176-181, 98 (2000).

[13] Thomsen E. G., Yang C.T., and Kobayashi S., **Mechanics of Plastic Deformation in Metal Processing**, Mc - Millan Co., New York, 1965.

[14] William R.D. Wilson, Steven R. Schmid, and Jiying Liu, "Advanced simulations for hot forging: heat transfer model for use with the finite element method", **Journal of Materials Processing Technology**, pp. 1912–1917, 155–156 (2004).

Modeling Virtual Cooperative Robots

Sanjay Kodiyalam[1], Duane Driggs[2], Amitava Jana[1]

Department of [1]Mechanical / [2]Electrical Engineering, Southern University, Baton Rouge, LA 70813
sanjaykodiyalam@engr.subr.edu, duanedriggs@engr.subr.edu, jana@engr.subr.edu

ABSTRACT

Motivated by the need to operate robots in environments not conducive for human presence a C/C++ project in Visual Studio is being developed for the display and control of Virtual Robots in the CAVE at Southern University's College of Engineering. In this CAVE-library based project the display is created using OpenGL: As a first case, of two robots operating in a master-slave configuration. The robots cooperatively move a solid bar that is jointly help by both. The links of the robot serving as master are moved by user input via a joystick. The slave follows the master automatically such that a dimensionless measure of the difference in the bar's position and orientation as defined by the two robots is minimized. The first case employs robots with three 2-axis links each. In this case the slave is able to follow the master instantaneously if the master is moved slowly otherwise there is a lag illustrating the need to improve the minimization algorithm.

Keywords: Cooperative, Robot, CAVE.

1. INTRODUCTION

Cooperating robots have been envisioned in areas not conducive for human presence: For example for target searching in a war zone [1]. For task planning of cooperating robots in similar circumstances, as in bomb disposal or handling radioactive materials with more than one robot required for increasing the load carrying capacity [2], a virtual reality based approach is being developed. While such task planning for actual robots involves kinematics, dynamics, and control [2, 3], virtual task planning primarily involves only kinematics with the cooperating robots forming closed kinematic loops. The current work stereoscopically visualizes two robots in a Cave Automatic Virtual Environment (CAVE) and targets virtual task planning with the end goal of driving real robots in the Mechatronics laboratory at Southern University.

2. CAVE-LIBRARY BASED PROJECT

A generic CAVE-library [4] based project to stereoscopically visualize multiple robots with multi-axis links has been developed [5]. In this project the fundamental constants of a particular robot configuration are the sizes of the links, the size of the bar to be held by the robots, the size of the robots' base, and the link angles that are fixed. The fundamental variables are the link angles that can be changed. A common data structure holds all of the constants and variables.

The CAVE library automatically spawns new synchronized threads, one for each display, in addition to the main thread that continues to operate asynchronously. The display threads also call a "navigation" routine which, in one of these threads, is for enabling navigation through the virtual world. By transforming the coordinate system of the virtual world relative to the fixed coordinate system of the CAVE the navigation routine currently provides for translations, rotations, and isotropic up/down scaling of the entire scene and eases user interaction with the virtual robots by helping in their global placement within the CAVE.

In the first case, two virtual robots are operated in a master-slave configuration with the master moved directly by user input while the slave attempts to follow automatically. In this case both master and slave robot movement routines are called by the display thread enabling navigation at the end of the navigation routine. To test against arbitrary positioning of a bar by the master robot, that is to be cooperatively held by the slave while avoiding possible problems in the convergence of its movement, both robots have, in this first case, three 2-axis links *i.e.* a total of six degrees of freedom (DOF) - equal to the DOF of a sold body in three dimensions. The project is primarily developed on a desktop with a single non-stereo perspective display called "CAVE simulator" and then tested in the CAVE with stereographic displays.

Display routine

As described previously [5] the geometry of the robotic links is first conceptualized in AUTOCAD and then constructed in the display routine using OpenGL. The matrix post-multiplication feature of OpenGL transformation accumulation makes it particularly simple to generate links of appropriate orientation: They are generated along a fixed (+*x*) axis in coordinate systems that are rotated and translated relative to the previous link by current link angles and link length respectively. While the data structure and other routines of the project allow for arbitrary multi-axis links the display routine is a not as general: It specializes to case of robots with two axis links (Fig.1) that can rotate around an axis along their length and also along the perpendicular shaft-axis. The display of the bar held by the robots is done twice - each following the transformations corresponding to a robot. Perfect cooperation between the two robots can then be evidenced by the display showing a single bar corresponding to overlap of the two displays of the bar.

Master movement routine

As described previously [5] the robot serving as the master is moved by a routine that accepts user input via a joystick. The display is augmented with a straight line emanating from the joystick towards its front (Fig. 1). On intersecting this line, any of the robots' links may be selected: The selected link identifies the master robot while the other becomes the slave. This link is highlighted in the display with a change in color (Fig. 1). Subsequent changes in the orientation of the joystick drive the changes in the orientation of the selected link – by changing all of its corresponding DOF. As the default frame rate of the display thread is high (~60 fps) changes in the orientation of the joystick between successive time frames is expected to be small

small and can therefore be computed as a vector $\delta\vec{\omega}$. The changes in the orientation defining triad of the joystick $\vec{J}_i$ are then given by:

$$\delta\vec{J}_i = \delta\vec{\omega} \times \vec{J}_i$$

which are solved for $\delta\vec{\omega}$:

$$\delta\vec{\omega} = \frac{1}{2}\sum_{i=1}^{3} \vec{J}_i \times \delta\vec{J}_i .$$

The projection of $\delta\vec{\omega}$ onto the axis i of the link l ($\vec{R}_i^l$) is the necessary incremental change in the link angle:

$$\delta\theta = \vec{R}_i^l \cdot \delta\vec{\omega} .$$

Again exploiting OpenGL's transformation accumulation $\vec{R}_i^l$ is read off as the i^{th} column vector in the rotation matrix contained in the accumulated affine transformation (see next subsection) that includes all transformations up to and including the link angle of interest. The affine transformation is constructed once in the master movement routine by following the sequence of transformations (rotations and translations) exactly as carried out in the display routine for the master robot. As the orientation relative to the robots changes upon navigation all vectors are always evaluated in the navigated or world frame of reference.

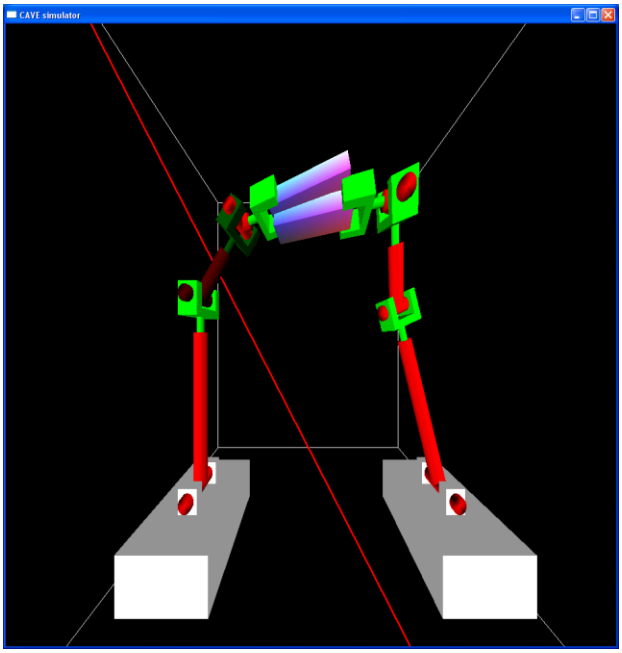


Fig. 1. Non-stereo perspective display of two robots in "CAVE simulator mode" on a desktop. Each has three 2-axis links. The second link of the left robot selected (therefore dull in color), via intersection of red line emanating from the joystick, making it the master while the other robot becomes the slave. Imperfect cooperation is illustrated to emphasize that the bar that is to be jointly held is displayed twice - corresponding one each to the positioning of the bar by one of the robots.

Slave movement routine

As described previously [5] the slave moves automatically in response to the master's movement such that the bar that is held by both robots is not strained. The movement of the slave is driven by a routine that iteratively minimizes a dimensionless measure Δ of the bar's strain. Δ is constructed from the difference between affine transformations [6] required for the positioning of the bar by the master (${}^{M}\boldsymbol{A}$) and slave (${}^{S}\boldsymbol{A}$) robots. With:

$${}^{M/S}\boldsymbol{A}_{4x4} = \begin{bmatrix} {}^{M/S}\boldsymbol{R}_{3x3} & {}^{M/S}\boldsymbol{t}_{3x1} \\ \boldsymbol{0}_{1x3} & \boldsymbol{1}_{1x1} \end{bmatrix}$$

($\boldsymbol{R}$ is the rotation and $\boldsymbol{t}$ is the translation of the bar's center – both relative to the coordinate system of the virtual world), Δ is chosen as:

$$\Delta = w_1 \frac{||{}^{S}\boldsymbol{t} - {}^{M}\boldsymbol{t}||}{(v_{bar})^{2/3}} + w_2 ||{}^{S}\boldsymbol{R} - {}^{M}\boldsymbol{R}||$$

where v_{bar} is the volume of the bar, and w_1 and w_2 are weight factors ($w_1 + w_2 = 1$) chosen to be equal as a first attempted case. For perfect cooperation $\Delta = 0$. Similar to the construction in the master movement routine the affine transformations ${}^{M}\boldsymbol{A}$ and ${}^{S}\boldsymbol{A}$ are constructed by following the sequence of transformations (rotations and translations) exactly as carried out in the display routine for the master and slave robots respectively. ${}^{M}\boldsymbol{A}$ is computed once while ${}^{S}\boldsymbol{A}$ is computed as many times as required in an iterative procedure that minimizes Δ . This procedure involves increasing or decreasing each of the slave's DOF as needed to reduce Δ and is an explicit search of all the slave's degrees of freedom that minimize Δ thereby avoiding the need to compute derivatives of Δ that will be required if other minimization algorithms such as the Newton-Raphson [7] are used.

As a first case the iterative procedure to minimize Δ sets the magnitude of change attempted specific to each DOF and is initially set to 5 degrees for all DOF. Subsequently it is reduced by a factor of 2 whenever both increment and decrement of the corresponding degree of freedom does not reduce Δ . The attempt to reduce Δ is done in sequence over all of the DOF and repeated until Δ is sufficiently small (10^{-5}) or till this procedure of minimization is deemed to have failed: When the magnitude of the attempted changes in the all the DOF are less than $5x10^{-8}$ degrees.

First (reference) case result

As described previously [5] in the case of both master and slave robots having six total DOF (three 2-axis links), and with their movement routines within the display thread, it is found that the slave is able to follow the master everywhere as long as the bar's position can be reached. Fig. 2 illustrates cases of the slave following the master. This cooperative movement happens "instantaneously"/interactively when the master is moved slowly – otherwise the movement of both robots "lag" the user input with differences in the bar's positioning by the two robots as evidenced by two non-overlapping displays for the bar. During this lag, as the master and slave movement routines are within the display thread, the display freezes. It results, in the CAVE, in a distortion of the perspective if the user's view is simultaneously changing. While the lag is acceptable visually as it does not lead to user frustration it illustrates that the algorithm for the slave's movement while being robust needs to be improved for better performance. It warrants a study of the magnitude of transient disagreement in the bars positioning by the two robots particularly because their motion is envisioned to eventually drive real robots in the Mechatronics laboratory.

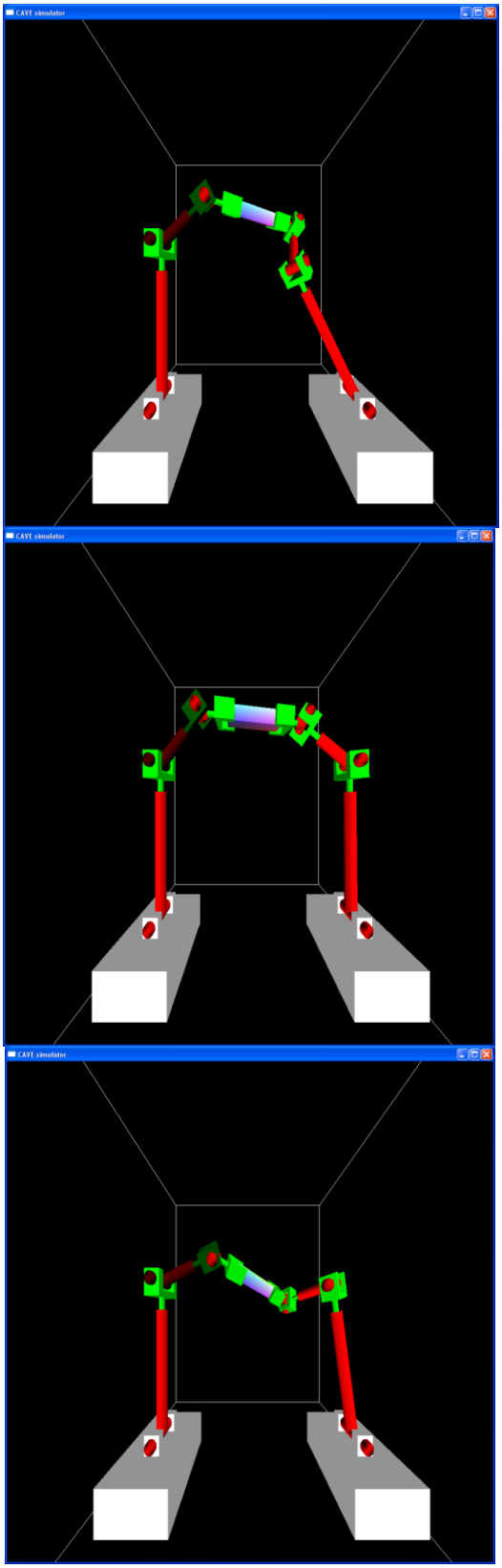


Fig. 2. Perspective views (CAVE simulator mode) of three cases in which the slave (right) robot follows the master (left) robot. Perfect cooperation is observed – otherwise two bars would be seen (as in Fig. 1). The 2nd link of master is being moved to generate the images.

3. PROJECT DEVELOPMENT

The virtual robot program is further developed in the following directions:

Separating robot movement and display operations
In order to avoid the freezing of the display routines both the master and slave have been moved to the main thread of the CAVE library. As this thread does not have an OpenGL context code for the matrix operations required by these routines has been developed. The execution speed of this code is noticeably slower than the equivalent OpenGL code in the display thread – as evidenced with the reference case configuration. Nevertheless the robot movement routines are operated in the main thread – with optimization of the matrix operations and the movement routines (particularly the slave) deferred to a later time.

Both robots as slaves to user controlled bar
For virtual task planning user control of the movement of the object cooperatively held by the robots is needed. This is

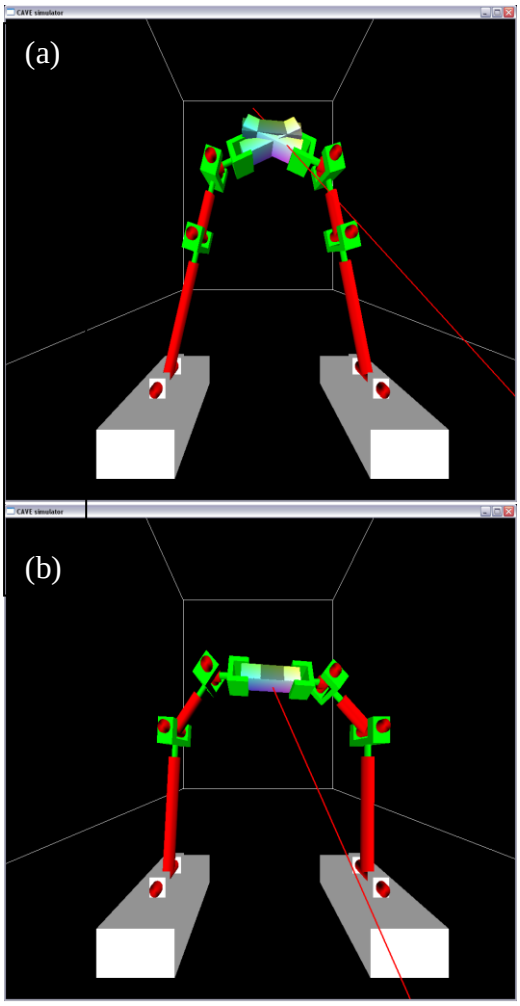


Fig. 3. Both reference case robots (each having 6 DOF) as slaves to the user controlled bar. (a) shows the user controlled bar in dull colors being outside of the workspace of both robots that are positioning their respective bars independently. (b) shows that when the user controlled bar moves into the robots' workspace there is perfect overlap of the bars: The robots have been able to follow the users' movement of the bar.

accomplished with both robots serving as slaves to an independent bar – that is introduced via a robot with no links. In order to translate this bar the project is modified to allow for any robot base to have translational DOF. The display correspondingly shows the selected base in a grey color and for the robot with no links the bar in dull colors (Fig. 3a). These DOF are updated via coupling to changes in the joystick position: As the position relative to the robots changes upon navigation all positions are always evaluated in the navigated or world frame of reference. Enabling these DOF for the robot with no links and the corresponding bar's rotations enables arbitrary positioning of the bar by the user. The reference case robots, each with 6 DOF, are able to follow the bar's movements as long as the bar is within their respective workspaces (Fig. 3b).

Tailoring to Laboratory robots: Convergence issues

As a first step in coupling the virtual and real worlds the virtual robots have been tailored to two identical TeachMover II Microbots [8] by having proportional link lengths [5] (Fig. 4 a & b). In this case the slave movement routine fails to converge to the users movement of the bar as each the robots have a total of only 5 DOF (Fig. 4a) – forcing the constraint of bar movement in a plane containing the 1st link. This convergence can be achieved by having the base of the robots be movable in a direction perpendicular to their initial separation (Fig. 4 c & d): A movable base will be required for real-world applications of involving movement over distances larger that in a fixed-robot workspace.

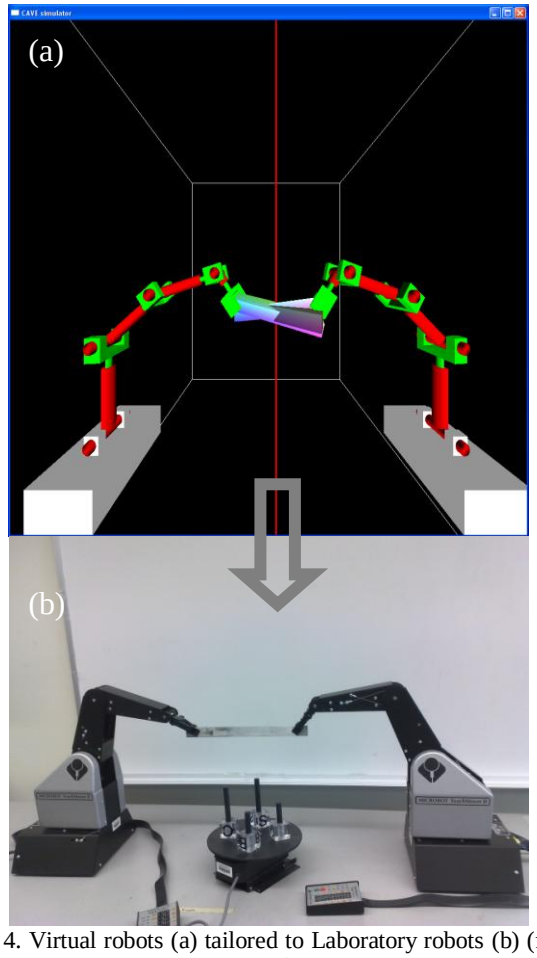


Fig. 4. Virtual robots (a) tailored to Laboratory robots (b) (from [5]). (a) shows that a rotation of the user controlled bar (dull colors) along the axis of the red line moves it out of the workspace of both the robots as they each have only five DOF: The 1st link can rotate around its axis, the 2nd and 3rd links around their respective support shafts, and the 4th link has both DOF.

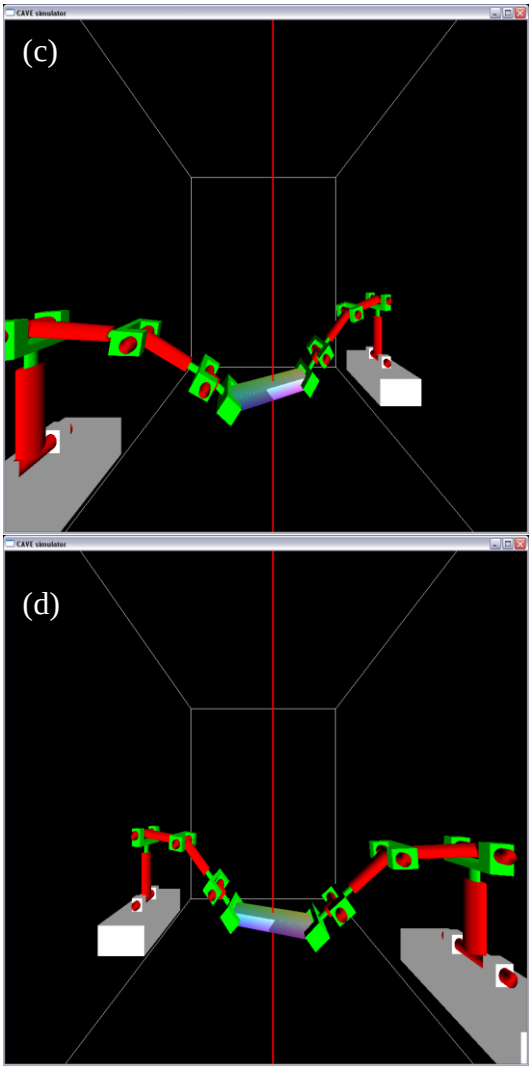


Fig. 4 (Contd) Convergence to the user controlled bar's movement (c & d) is made possible by having the robots in Fig. 4a have their base moveable in a direction perpendicular to their initial separation.

4. CAVE AT SOUTHERN UNIVERSITY

The CAVE at Southern's College of Engineering is a projection based Virtual Reality (VR) system (Fig. 5): An 8ft-cubed space with three screen-walls and the floor serving as displays (Fig. 5). The images are projected via mirrors used to set the required optical distance between the projectors and screens/floor under the constraint of the room space available to house the CAVE. Active stereographic viewing is enabled with separate projections for the left and right eyes synchronized with the eyeware. The position and orientation, *i.e.* 6-DOF, of two sensors is continuously tracked in the CAVE. These DOF are relative to the fixed coordinate system of the CAVE that has the origin at the center of the floor with the x-axis towards the right

screen, the y-axis towards the top and the z-axis away from the front screen. One sensor is attached to the eyeware and the other that is attached to a joystick. The perspective transformations required for the displays is automatically generated and applied by the CAVE library [4] based on the eyeware DOF while the joystick DOF is made available to the user application to enable navigation and any other interaction with the virtual world.

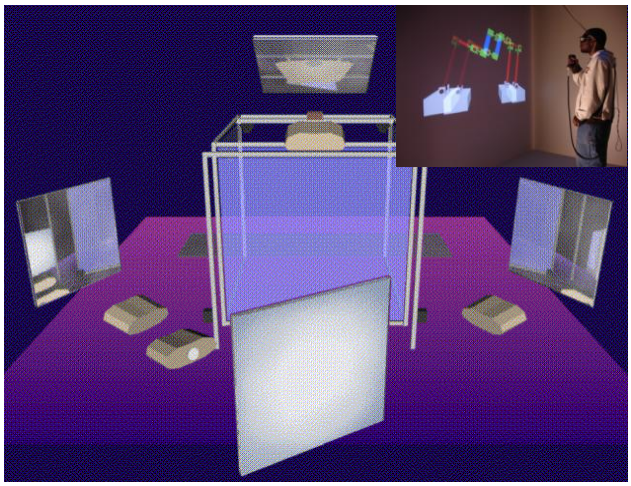


Fig. 5. Schematic [9] of the CAVE at Southern showing the projectors, mirrors, and screens. The inset (from [5]) shows one of the authors (Driggs) running the project in the CAVE at Southern.

5. ONGOING WORK: CONNECTING WITH LABORATORY ROBOTS

A laboratory Scorbot [10] has been interfaced with a computer [5]. The Scorbot's software has been used to drive its motion: the gripper follows various profiles in 3D space defined by a user-program via keyboard input. A profile has been copied from the Scorbot's controller to the hard disk, edited by the user, and then used to re-drive the Scorbot. The ability to use the modified profile suggests that output from the virtual robots program may be used to drive laboratory robots – after overcoming problems with having only 5 DOF and the inability to directly control the angular orientation of the laboratory robot's gripper. The first attempt will be file-based: Storing the virtual robots' movements in a file to drive the laboratory robots. Subsequently dynamic coupling will be attempted with the virtual and real worlds linked over using network-based software analogous to that proposed in remote laboratory experiments [11]. For the eventual coupling to the virtual robots, as illustrated in Fig. 4, four robots in the Mechatronics laboratory have been refurbished and upgraded to TeachMover II Microbots [9].

6. CONCLUSIONS

The VR environment of the CAVE provides a novel approach to task planning of cooperating robotic systems. Coupling the virtual robots to and real laboratory robots is expected to reveal the advantages of this approach.

7. ACKNOWLEDGEMENT

This work is supported by LA-SiGMA - an NSF-EPSCoR grant (NSF EPS-1003897.). The Center for Energy and Environmental Studies at Southern University is acknowledged for supporting hardware upgrades to the CAVE.

8. REFERENCES

[1] Adam Ray, "Cooperative Robotics Using Wireless Communication," Master's Thesis, Auburn University, (2005).

[2] A. Jana, D. M. Auslander, R. N. Dave, S. S. Chehl, "Task Planning for Cooperating Robot Systems," *Proceedings of the First World Automation Congress*, Maui, Hawaii, (1994).

[3] H. P. Huang, and R. S. Chen, "Modelling and Adaptive Coordination control of a Two-robot System," *Journal of Robotic Systems*, **9** (1), 65-92, (1992).

[4] Dave Pape, Carolina Cruz-Neira, Marek Czernuszenko, *CAVE Library version 2.6*, Electronic Visualization Laboratory, University of Illinois at Chicago, (1997). Currently used: Version 3.2, CAVELib™, Mechdyne Corporation, Marshalltown, Iowa.

[5] D. Driggs, S. Baylor, B. M. S. Kuppala, S. Kodiyalam, K. Sekeroglu, A. Jana, "From Virtual to Laboratory Robots," *Proceedings of the 2011 ASEE-GSW Conference*, March 2011, Houston, Louisiana, Session **FC1-1**.

[6] http://en.wikipedia.org/wiki/Affine_transformation

[7] http://en.wikipedia.org/wiki/Newton's_method

[8] D. Eavey, "Robotic Workcells and System Interfacing," *UMI-MICROBOT*, **1**, UMI-MICROBOT Inc., Sunnyvale, CA, (1988). Currently used: Teachmover II, Questech Inc,

[9] Figure from http://www.cs.uic.edu/~kenyon/Conferences/GILKY/CAVE_DOD.html

[10] *SCORBOT*, Intelitek Inc., Manchester, New Hampshire.

[11] S. Paluri, K. Sekeroglu, M. H. Yilmaz, A. Jana, "Web-based Experimentation for Efficient Learning," *Proceedings of the 2011 ASEE-GSW Conference*, March 2011, Houston, Louisiana, Session **FC1-1**, Pgs:1-6.

Application of Artificial Neural Networks for Analysis of Misalignment in Rotating Machinery Using Coast Down Time

G. R. Rameshkumar
Department of Mechanical & Industrial Engineering, Caledonian College of Engineering,
P.O. Box 2322, CPO 111 Seeb, Sultanate of Oman

Dr. B. V. A Rao
VIT University,
Vellore 632 014, Tamil Nadu, India

and

Dr. K. P. Ramachandran
Caledonian College of Engineering,
P.O. Box 2322, CPO 111 Seeb, Sultanate of Oman

ABSTRACT

Mechanical fault such as shaft misalignment is one of the major concerns in rotating machinery. The behaviour of rotating machinery during deceleration period when the power supply to the drive is stopped is known as Coast Down Phenomenon. The total time elapsed for the entire momentum due to sustained operation of the rotating machinery to dissipate is termed as Coast Down Time. The Coast Down Time is used as a diagnostic condition monitoring parameter to understand and detect shaft misalignment. Experiments were carried out on Forward Curved Centrifugal Blower Test setup to record the CDTs at selected blower shaft cut-off speeds of 1000 rpm, 1500 rpm, 2000 rpm and 2500 rpm respectively for different levels of combined horizontal and vertical parallel misalignment and at various orders of combined horizontal parallel and angular misalignment conditions. These sets of experimental CDTs were used to train and test the neural network. This paper outlines the implementation of Radial Basis Function neural network, a new generation of ANNs for analysis and prediction of CDTs for different combined misalignment fault conditions. RBF has been found to be successful in the prediction of CDTs and the results showed that the ANN predicted CDT values are very close to the experimental CDT values for various combined misalignment fault conditions.

1. INTRODUCTION

Mechanical fault such as shaft misalignment is the most common causes of vibration in rotating machinery and one of the major concerns in modern industry. Misalignment in shafts leads to increase in vibration, and an increase in radial and axial loads, thereby absorbing more energy and power. This, in turn, causes premature wear and even catastrophic failure of bearings, seals, coupling and other components in the machinery [1]. In actual operating environments in industries, achieving perfectly aligned condition is very difficult and a minute amount of misalignment is always present.

When the power supply to any rotating system is cut-off, the system begins to lose its momentum gained during sustained operation and finally comes to a halt. The behaviour of the rotor system during this period is known as the Coast Down Phenomenon (CDP). The exact time period between the power cut-off time and the time at which the rotor stops is called Coast Down Time (CDT) [2]. CDT is the total time taken by the system to dissipate the momentum acquired during sustained operation. The performance of the misaligned cylindrical [3] and three-lobe journal bearings [4] were evaluated. It was found that the CDT decreases with the increase in misalignment. This is because both the bearing friction and the power loss increase with increase in misalignment. Increase in misalignment will cause an increase in coefficient of friction and reduces the film thickness resulting in an increase in damping factor.

Artificial neural network (ANN) is a representation of the computational architecture of the human brain. It is an established tool for effortless computation and its application in the area of automated fault detection and diagnosis of machine condition is very promising [5]-[8]. Radial basis function neural networks (RBFNN) require lesser neurons than the standard feed forward back-propagation networks (FFBPN) [9]-[11]. They can be trained in a fraction of time. RBFNN is used to model engineering systems and found that it is efficient, reliable and robust technique [12].

The CDT can be used as one of the diagnostic parameter to assess the condition of the rotating machinery [13]. A summary of literature reveals that the application of ANN is not extended to CDT prediction till now. This paper proposes a neural network model that will be able to predict the CDTs for different combined misalignment fault conditions.

2. EXPERIMENTAL TEST SETUP AND PROCEDURE

The photographic view of the Forward Curved Centrifugal Blower Test Setup [14] used for this investigation is shown in Figure 1. A Forward Curved (FC) centrifugal blower is mounted on a shaft with length 315 mm and diameter of 20 mm at the center position of 190 mm between two anti-friction bearings. The blower shaft is connected through an electromagnetic coupling to a variable speed DC motor. Two inductive proximity sensors are used to measure the speeds of blower and motor independently. An instrumentation control panel is built-

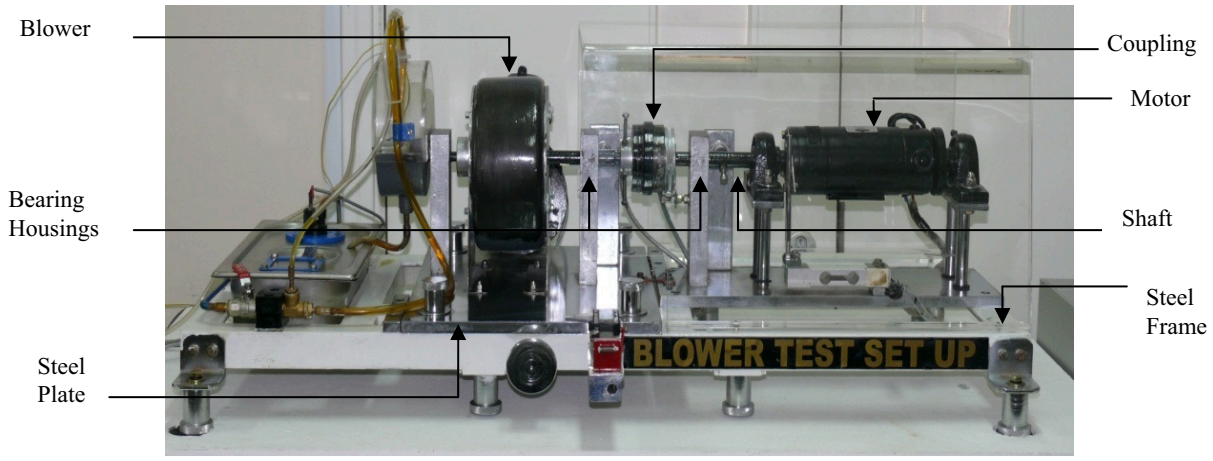


Figure 1: Centrifugal Blower Experimental Test Setup

in display and control the variables. A software developed using Visual Basic is used along with instrumentation to control the operation of the test setup and to record the motor and blower speeds as well as CDT for each test run for different shaft cut-off speeds. The software has an ability to record CDT with an accuracy of 60 ms intervals, and corresponding deceleration speed of blower shaft at each test run.

Initially, the coast down time for each test run at various cut-off speeds were recorded and used thereafter as baseline reference for further investigation, analysis and comparison. The baseline CDTs obtained for blower shaft at cut-off speeds of 1000 rpm, 1500 rpm, 2000 rpm, and 2500 rpm are 1560 ms, 2040 ms, 2580 ms, and 3000 ms respectively under healthy operating conditions.

Combination of horizontal and vertical parallel misalignment conditions was created between the blower shaft and motor shaft. Three levels of combined offset distances of 0.10 mm, 0.20 mm and 0.30 mm respectively have been introduced at the blower shaft end.

Combination of horizontal parallel misalignment and angular misalignment conditions was created between the blower shaft and motor shaft. Three levels of combined misalignment of (0.15 mm & 0.046^0), (0.25 mm & 0.076^0) and (0.35 mm & 0.107^0) respectively have been introduced.

Experimental tests were conducted at blower shaft cut-off speeds of 1000 rpm, 1500 rpm, 2000 rpm, and 2500 rpm respectively. The coast down time and the respective deceleration speeds of blower shaft were recorded. Each experiment was repeated 5 times for consistency under the same operating conditions to record CDTs. To verify the validity and accuracy of CDT data error analysis was carried out for sample size of 5, it was ascertained that the magnitude of the error is less than 1 percent in all cases of both healthy and fault conditions. The repeatability of CDT data is around 99 percent. These experimental data [15] are used to train an ANN to predict the CDTs for various combined misalignment fault conditions.

3. RADIAL BASIS FUNCTION NEURAL MODEL

A Radial Basis Function (RBF) is a two-layer network whose output units form a linear combination of the basis functions computed by the hidden units. A function is radially symmetric (or is an RBF) if its output depends on the distance of the input sample (vector) from another stored vector. Neural networks whose mode functions are radially symmetric are referred to as radial basis function networks [16].

The general model of RBF neuron is shown in Figure 2. The transfer function for a radial basis neuron is radbas. The radial basis neuron receives as net input, the vector distance between its weight vector w and the input vector p, multiplied by the bias b. The basis functions in the hidden layer produce a localized response to the inputs so that each hidden unit has a localized receptive field. The basis function can be viewed as the activation function in the hidden layer. The outputs of the hidden unit lie between 0 and 1. The closer the input to center of the Gaussian, the larger the response of the node. The node produces an identical output for inputs with equal distance from the center of the Gaussian; it is called a radial basis. The output unit finds a linear combination of nonlinear basis functions and thus the network performs a nonlinear transformation of the input.

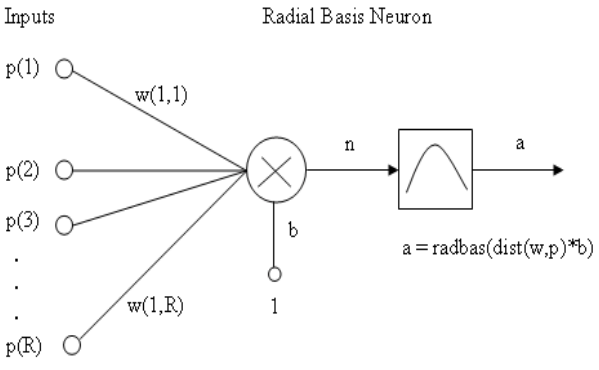


Figure 2: Radial Basis Function Neuron Model

RBF network is capable of approximating any arbitrary mapping. The main difference between the RBF network and the back-propagation network is in their basis functions. The radial basis function covers only small regions, whereas the sigmoid function used in neural network assumes nonzero values over an infinitely large region of the input space. Classification tasks are more amenable to the RBF network than the back-propagation network in the case when the problem is extended to higher dimensions [17].

4. RBF TRANING PROCEDURE

The radial basis neural networks have been designed by using the function newrb in the neural network toolbox supported by MATLAB [18]. The function newrb iteratively creates a radial basis network by including one neuron at a time. Neurons are added to the network until the sum of squared error is found to be very small or the maximum numbers of neurons are reached. At each iteration, the input vector, which will result in lowering the network error most, is used to create a radial basis neuron. During the training, each of the connecting weights of the individual neuron is compared with the input signals. The distance between the connecting weights determines the output of hidden neurons and input vector, which is further multiplied by bias. Bias is an additional scalar quantity being added between the neuron and fictitious neuron. The output is propagated in a feed forward direction to the output layer neuron, which will give the output if the connection weights are close to the input signal. This output is compared with target vector. If the error reaches the error goal, then training is terminated, otherwise the next neuron will be added. The connecting weights are modified each time by changing the maximum neurons and the spread constant. The value of maximum neuron and spread constant keeps on changing till the network is trained properly. Radial basis networks can be used to approximate functions; newrb adds neurons to the hidden layer of a radial basis network until it meets the specified mean squared error goal.

5. RESULTS AND DISCUSSIONS

The input parameters were normalized before being applied to train and test the networks. The CDTs were normalized in the range from 0 to 1. Similarly, deceleration speeds were normalized in the range from 1 to 0. In the present work Radial basis function neural network is considered for predicting the CDT.

5.1 Combined Parallel Misalignment

The RBF network was trained for the various levels of combined parallel misalignment conditions. From 27 CDT data points at various level of combined parallel misalignment, 21 data were used for training and the remaining 6 data were used for testing at blower shaft cut-off speed of 1000 rpm. From 32 CDT data

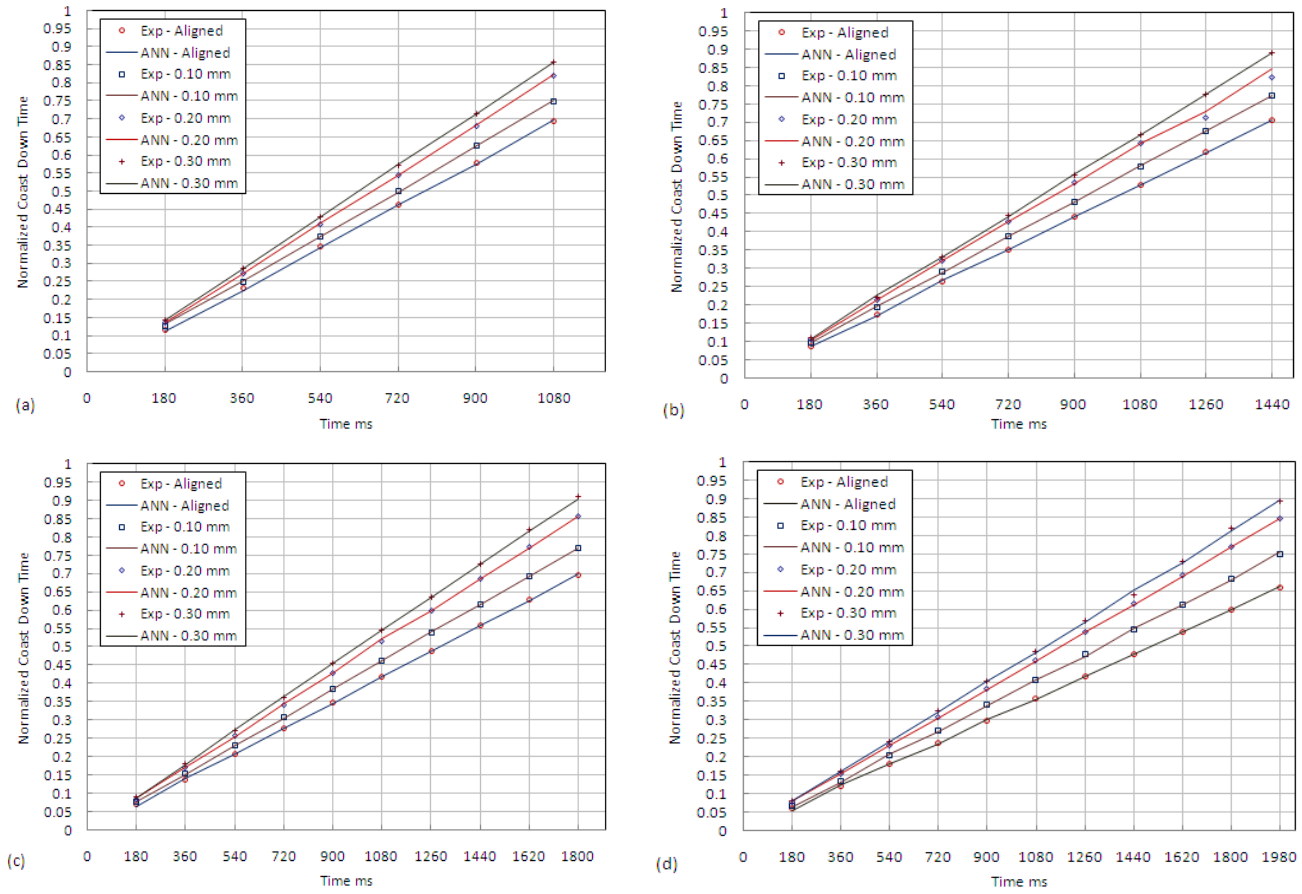


Figure 3: Experimental and Artificial Neural Networks predicted CDTs at blower shaft cut-off speeds of a) 1000 rpm, b) 1500 rpm, c) 2000 rpm and d) 2500 rpm for various combined horizontal and vertical parallel misalignment conditions

points, 24 data were used for training and the remaining 8 data were used for testing at 1500 rpm. From 40 CDT data points, 30 data were used for training and the remaining 10 data were used for testing at 2000 rpm. From 44 data CDT data points, 33 data were used for training and the remaining 11 data were used for testing at 2500 rpm. Variation in number of data points for training and testing is due to higher CDT values at higher cut-off speeds.

The trained network was used to predict the CDTs at different cut-off speeds, at various offset distances in both the direction, and the results were compared with the experimental and ANN predicted values

5.2 Combined Parallel and Angular Misalignment

The RBF network was trained for the various orders of combined horizontal parallel and angular misalignment conditions with the same number of sets of data points used in combined parallel misalignment were used for training and testing under same shaft cut-off speeds, and the results were compared with the experimental values as shown in Figure 4.

The neural network predicted CDTs were compared with the experimental CDT values at the coast down time intervals of 180-2160 ms. It has been observed that artificial neural network modeling of system based parameters are found to match closely with the experimental data at aligned and at various combined misalignment conditions.

By comparing the ANN predicted CDT values with experimental CDT values; artificial neural networks modeling of system is found to be satisfactory for predicting the CDTs for combined parallel misalignment and combined parallel and angular misalignment faults.

The absolute standard deviation and root mean square error calculated for each fault at various cut-off speeds are tabulated in Table. 1. From these results, for all the two cases, it has been observed that the error values as described by ABSD, RMSE are well within the permissible limits. This proves the efficiency of the methodology adapted in this research. Hence the ANN based prediction of CDT for mechanical fault in rotating machinery has been found successful and reliable. The artificial neural network modeling to predict coast down time is highly justified for the CDT analysis as one of the diagnostic condition monitoring parameter to assess the condition of the rotating machinery.

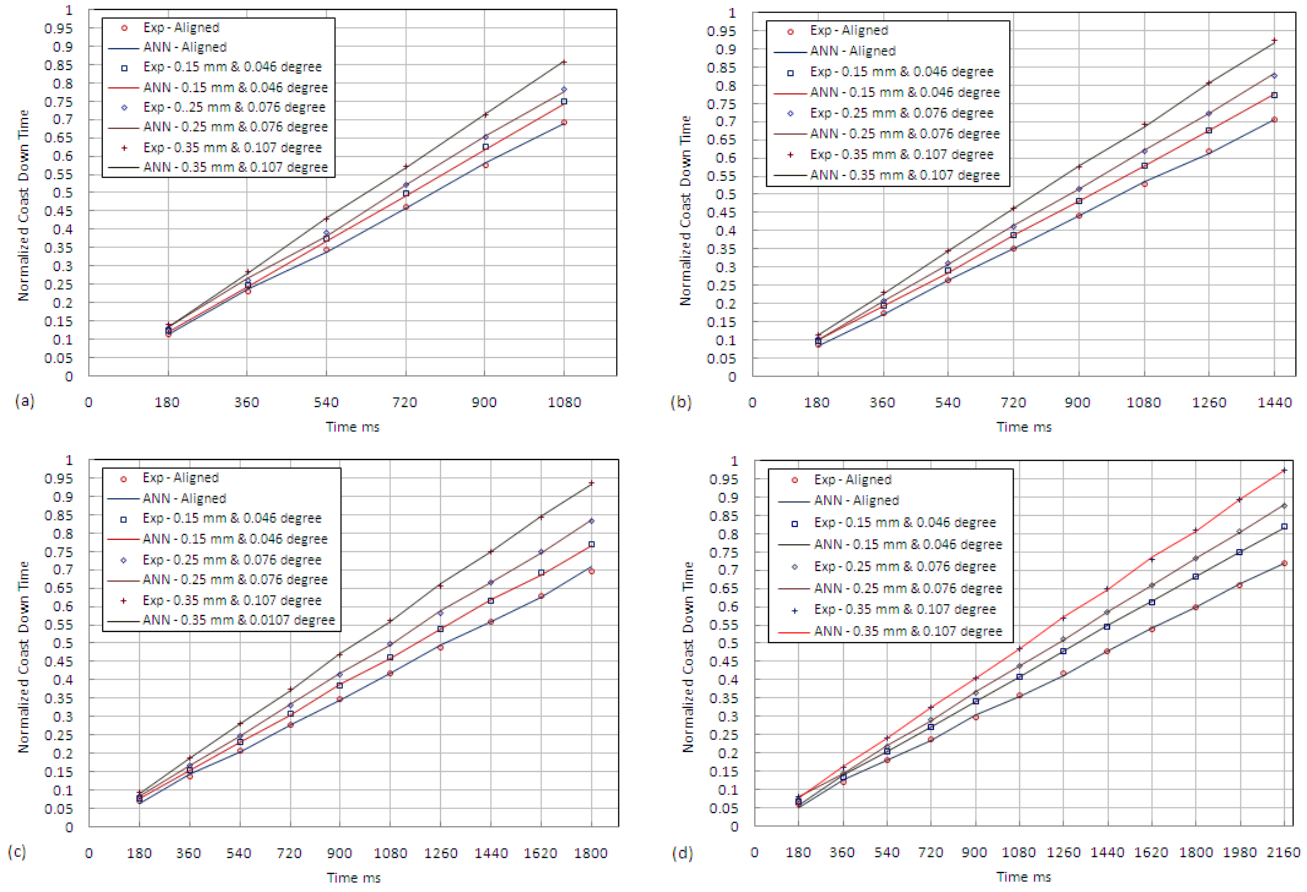


Figure 4: Experimental and Artificial Neural Networks predicted CDTs at blower shaft cut-off speeds of a) 1000 rpm, b) 1500 rpm, c) 2000 rpm and d) 2500 rpm for various combined horizontal parallel and angular misalignment conditions

TABLE I

ABSD and RMSE values at different cut-off speeds for various orders of misalignment fault conditions

Mechanical faults	Cutoff Speeds rpm							
	1000 rpm		1500 rpm		2000 rpm		2500 rpm	
	ABSD	RMSE	ABSD	RMSE	ABSD	RMSE	ABSD	RMSE
Combined Horizontal and Vertical Parallel Misalignment	0.00243	0.01491	0.00284	0.01287	0.00186	0.01462	0.00213	0.01984
Combined Horizontal Parallel and Angular Misalignment	0.00484	0.01934	0.00229	0.01157	0.00249	0.01223	0.00259	0.03747

ABSD = Absolute Standard Deviation, RMSE = Root Mean Square Error

6. CONCLUSIONS

Experimental evidences are showing CDT as an indicator to detect shaft misalignment fault. The complexity in identifying parallel or angular misalignment fault exactly, the need of automatic fault diagnosis using ANNs tool as a viable solution is explored in this paper. The radial basis network approach has been used for CDT data prediction as it employs limited neurons for the construction and requires lesser computational time in modeling the system. The proposed technique of using radial basis function requires only limited experimental data points to train, model and predict the system behaviour. RBF network was successfully implemented to predict CDTs for various orders of combined parallel and angular misalignment mechanical fault conditions. The performance of the RBF network in predicting the CDT is found to be more accurate.

ACKNOWLEDGMENT

The author[1] thankfully acknowledges the generous and continuous support of the Management, Principal and Dean of the College for carrying out this research work for his PhD under Staff Development.

APPENDIX A

Absolute Standard Deviation:

$$\mathrm{ABSD} = \frac{\sum |\,(\text{ANN value} - \text{Experimental value})\,|}{\text{Number of data points}}$$

Root Mean Square Error:

$$\mathrm{RMSE} = \sqrt{\frac{\sum \left(\dfrac{\text{Experimental value} - \text{ANN value}}{\text{Experimental value}} \right)^2}{\text{Number of data points}}}$$

REFERENCES

[1] Piotrowski John, **Shaft Alignment Handbook**, 3rd edition, CRC Press; Taylor & Francis Group LLC; New York, 2007.

[2] T. L. Daugherty and R. J. Craig, "Coast down time as a mechanical condition indicator for vertical axis motors with grease-lubricated ball bearings," **American Society of Lubrication Engineers Transactions**, 1977, pp. 349-357.

[3] B. S. Prabhu, "An experimental investigation on the misalignment effects in journal bearings," **Tribology Transactions**, vol.40, no. 2, 1997, pp. 235-242.

[4] P. Arumugam, S. Swarnamani, and B. S. Prabhu, "Effects of journal misalignment on the performance characteristics of three-lobe bearings," **Wear** 206, 1997, pp. 122-129.

[5] K. Mehrotra, C.K. Mohan and S. Ranka, **Elements of artificial neural networks**, Mumbai, Penram International, 1997.

[6] A. K. Nandi, "Advanced digital vibration signal processing for condition monitoring," **Proc. 13th International congress and exhibition on condition monitoring and diagnostic engineering management (COMADEM2000),** Houston, Tex, USA, 2000, pp. 129-143.

[7] R. B. Randall, Ed, "Special issue on gear and bearing diagnostics," **Mechanical Systems and Signal Processing**, vol. 15(5), 2001, pp. 827-1029.

[8] B. A. Paya, I. L. Esat, and M. N. M. Badi, "Artificial neural network based fault diagnosis of rotating machinery using wavelet transforms as a preprocessor," **Mechanical Systems and Signal Processing**, vol. 11(5), 1997, pp. 751-765.

[9] A. Baraldi, and N. A. Borghese, "Learning from data: general issues and special applications of radial basis function networks," **Tech. Rep. TR-98-028, International Computer Science Institute**, Berkeley, California, USA, 1998.

[10] J. Park, and I. W. Sandberg, "Universal approximation using radial basis function networks," **Neural Computation**, vol. 5, no.2, 1993, pp. 305-316.

[11] L. B. Jack, and A. K. Nandi, "Feature selection for ANNs using Genetic Algorithms in condition monitoring," **European Symposium on Artificial Neural Networks**, Bruges (Belgium), 1999, pp. 313-318.

[12] L. Govindarajan, and P. L. Sabarathinam, "Prediction of vapor-liquid equilibrium data by using radial basis neural

networks," **Chem. Biochem**. Eng. Q. 20 (3), 2006, pp. 319-323.

[13] G. R. Rameshkumar, B. V. A. Rao, and K. P. Ramachandran, "An experimental investigation to study the effect of misalignment using CDT as a condition monitoring parameter for rotating machinery," **22nd International Congress, COMADEM 2009**, San Sebastian, Spain, 2009, pp. 531-539.

[14] G. R. Rameshkumar, B. V. A. Rao, and K. P. Ramachandran, "Coast down time analysis used as a condition monitoring parameter to experimentally assess the effect of misalignment for forward curved centrifugal blower," **Proc. Of the International Conference on Advances in Mechanical Engineering**, SVNIT, Surat, India, 2009, pp. 121-125.

[15] G. R. Rameshkumar, B. V. A. Rao, and K. P. Ramachandran, "Condition monitoring of forward curved centrifugal blower using coast down time analysis," **International Journal of Rotating Machinery**, Vol.2010, Article ID 962804, 12 pages, doi:10.1155/2010/962804.

[16] F. Li. Min, **Neural networks in computer intelligence**, 1st ed. Singapore: McGraw-Hill, 1994.

[17] L. Govindarajan, **Optimal design of reactors**, PhD Thesis, Annamalai University, India, 2005.

[18] Neural Networks Toolbox **User's guide**. 1sr ed. The math Work Inc, Mass, 1994.

New Developments in Ultrasonic-assisted Grinding and Dressing

Taghi Tawakoli
Institute of Grinding and Precision Technology, KSF, Furtwangen University,
Villingen-Schwenningen 78054, Germany

Abdolreza Rasifard
Institute of Grinding and Precision Technology, KSF, Furtwangen University,
Villingen-Schwenningen 78054, Germany

and

Bahman Azarhoushang
Institute of Grinding and Precision Technology, KSF, Furtwangen University,
Villingen-Schwenningen 78054, Germany

ABSTRACT

The use of the ultrasonic assistance results in significant improvements in machining processes. Nevertheless, its use in the grinding and dressing processes has still not been widely introduced into industrial environments. This is mostly due to the difficulties in achievement of an efficient superimposing of the vibrations on the industrial workpiece or tool in grinding and to the grinding wheel or the dressing tool in the dressing process. This paper introduces a recent developed ultrasonic vibration system with an innovative blocksonotrode which can vibrate industrial workpieces of different sizes and weights during grinding. An innovative ultrasonic spindle is also introduced which facilitates the ultrasonic-assisted dressing of grinding wheels with form rollers. The experimental investigations carried out at the KSF Institute show that the new developed ultrasonic systems can improve the process and increase the productivity.

Keywords: Ultrasonic Assistance, Grinding, Dressing.

1. INTRODUCTION

In ultrasonic-assisted machining, the kinematics of the conventional machining process are superimposed with an additional vibration in the micrometre range at frequencies in the ultrasonic range, as a result of which a considerable improvement in the process, such as reduction in active machining forces, improvement in the supply of coolant into the process zone, reduction in tool wear, etc. can be achieved [1-4].

In the ultrasonic assisted machining, the ultrasonic vibrations are produced by means of an ultrasonic vibration system which converts electrical energy to mechanical vibrations and transfers the vibrations through the workpiece or tool into the process zone. This system generally consists of a transducer, booster, and sonotrode (Fig. 1).

In the conventional ultrasonic vibration systems, the tool or workpiece which is directly attached to the sonotrode must either be very small or operate with a special construction as a part of the vibration chain. These all limit the efficiency of ultrasonic assistance in the grinding process. Thus, the lack of a system which is able to vibrate industrial workpieces of different sizes and weights is felt. For this purpose, an inexpensive innovative vibration system has been developed.

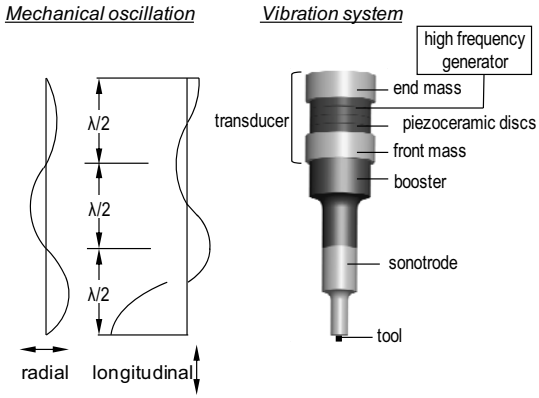


Fig.1 Conventional ultrasonic vibration system [5]

Research reports also indicate that remarkable improvements can be achieved in dressing with ultrasonic assistance, particularly in the dressing of superabrasive CBN and diamond grinding wheels [6-9]. Nevertheless, the use of ultrasonic assistance in dressing with diamond form rollers despite of their widespread industrial use has not yet been realized for technical and economic reasons.

2. NEW ULTRASONIC SYSTEM WITH BLOCK SONOTRODE

In order to vibrate workpieces with different shapes, dimensions and weights an innovative ultrasonic system was developed in this work. In this system, the block sonotrode with through holes can be account as the key part amongst the others. It is a special designed plate with a multi-resonant frequency response. Fig. 2 shows the new ultrasonic system with its resonance vibration at 20 kHz simulated using finite elements method. The multi-resonance behavior of the block sonotrode can only be achieved when the positions and diameters of the through holes are adequately designed. In addition, the block sonotrode serves as the workpiece clamping plate and therefore should possess a certain rigidity to ensure an accurate process.

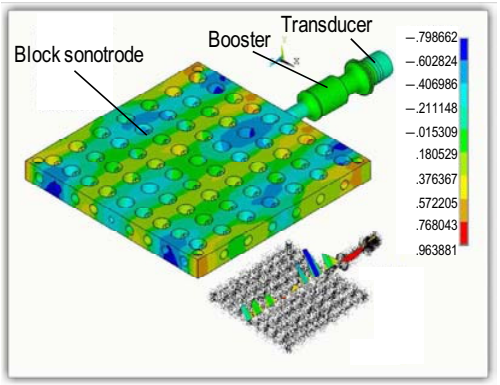


Fig. 2 Resonance simulation of the new ultrasonic system at 20 kHz

3. ULTRASONIC DRESSING SPINDLE

As mentioned, the use of ultrasonic assistance in dressing with diamond rollers is of great importance regarding the widespread use of these dressers. However, this subject has not yet been realized because of the difficulties caused by the dimensions and the weight of the roller dressers and dressing spindles, which are normally used in engineering applications. In the following, a new innovative ultrasonic spindle will be introduced that facilitates the ultrasonic-assisted dressing with form rollers.

From kinematical point of view, ultrasonic assisted dressing with form rollers can be achieved theoretically either by a form roller on which ultrasonic vibrations are applied in the direction of the grinding wheel, or by a form roller which in the radial direction is alternately stretched and compressed in the micrometer range at a frequency in the ultrasonic range. The developed ultrasonic spindle works after the second principle. So it generates beside the rotational movement the radial vibration in a form roller.

The design of the ultrasonic dressing spindle is shown schematically in Fig. 3. The vibration system consists of a piezoceramic acoustic converter and a sonotrode that functions also as the shaft of the spindle. By resonance standing longitudinal waves with fixed vibration, antinodes and nodes are produced in the vibration chain. On the other hand, the standing longitudinal wave leads to a periodic enlargement and reduction of the waveguide cross-section (s. also Fig 1). The cross-section change is greatest in the nodal points and becomes zero on the longitudinal vibration antinodes, where no local longitudinal extensions occur. The size of the change in cross-section depends on the longitudinal vibration amplitude, cross-section diameter, and Poisson's ratio. As the diameter of the sonotrode increases, the radial amplitude increases. The form roller, coated with diamond on the face is located at a nodal point of the sonotrode. Therefore, the periodic change in cross-section of the vibration node leads to transmission of the radial uniform vibrations into the form roller. This radial vibration is amplified by the special shape of the dressing roller [10].

The engagement paths of the diamonds of the form roller are influenced by the radial vibration of the form roller. This leads to the formation of different wheel topographies and consequently different grinding results with conventional and ultrasonic assisted dressing [5].

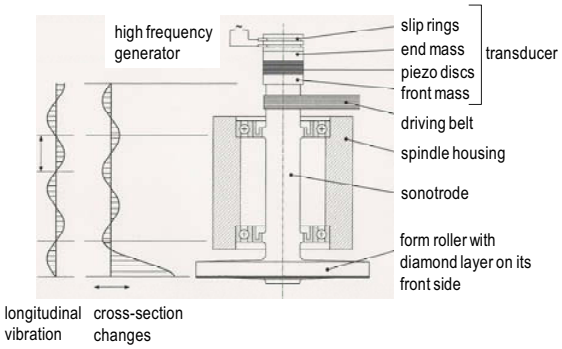


Fig. 3 Construction of the new developed ultrasonic dressing spindle [10]

4. EXPERIMENTAL INVESTIGATION WITH NEW ULTRASONIC SYSTEM

The experiments were carried out using a flat vitrified bond CBN wheel and a Nimonic (80A) workpiece that is known as a difficult to cut material. Fig.4 illustrates the experimental setup. The grinding forces are measured by a dynamometer which is mounted beneath the workpiece and on the machine table. The grinding wheel wear was measured by determining the wheel periphery distance from a reference point on the flange. This was conducted by means of a sensor measurement system, Z-Nano HP by NovoBlum GmbH, which was integrated in the grinding machine [11]. The grinding temperature was measured using a thermocouple that was mounted under the ground surface.

The experimental equipment consists of the following:

- Machine tool: Elb Micro-Cut AC8 CNC universal surface grinding machine
- Surface roughness and profile tester: Hommel-Werke model T-8000
- Dynamometer: Kistler piezoelectric dynamometer model 9255B
- Eddy current displacement measurement system: Micro epsilon eddyNCDT 3300, to measure the amplitude of vibration
- Ultrasonic Vibration Generator: AMMM-1000 W generator-MP Interconsulting, Frequency ranges of 18.000 to 50 kHz
- Thermocouple type K (NiCr/NiAl) after DIN EN 60584-1 with an outer diameter of 0.5 mm

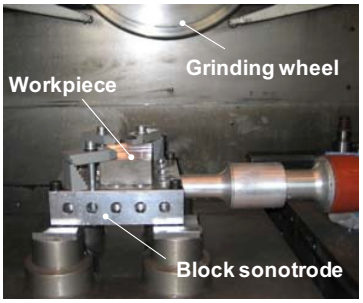


Fig. 4 Experimental set-up

Table 1 Major machining parameters in investigation with new ultrasonic system

Grinding wheel	Vitrified bond CBN grinding wheel, B126 C150
Workpiece	Nimonic A80
Grinding conditions	Up-grinding Feed speed v_{ft} = 2000 mm/min; Cutting speed v_c = 90 m/s; Depth of cut a_e = 0.5 mm
Coolant	Grinding oil, 100 lit/min
Grinding process	Surface grinding
Dresser	Electroplated diamond form roller D427
Dressing conditions	Overlap ratio U_d = 3; Wheel peripheral speed v_{sd} = 60 m/s; Depth of dressing cut a_{ed} = 5 x3 µm, Dressing speed ratio q_d = 0.5
Ultrasonic vibration	Parallel to the feed direction, Frequency f_{US}=20.5 kHz, Amplitude A_{US} = 10 µm

The settings of main machining parameters for the present study are summarized in Table 1.

Fig. 5 shows the specific grinding forces (F'_t and F'_n) when ultrasonic assisted grinding (UAG) and conventional grinding (CG), with increasing material removal V'_w. Fig. 6 shows the workpiece temperature when UAG and CG. It is obvious that with ultrasonic assistance the grinding forces and workpiece temperature are lower. This may be due to several reasons. First, in machining of tough nickel-based superalloys like Nimonic, the chip formation is improved by means of ultrasonic vibration. Due to the high-frequency interaction of active grains on the workpiece, the cutting process in ultrasonic assisted grinding process becomes discontinuous and ultrasonic impact action occurs, causing the material to begin to rollover more easily as well as more micro-cracking propagation in the cutting zone which both make an effective interaction between grits and the workpiece surface [12]. Therefore, the grinding forces and frictional effects are decreased, so that less plastic deformation occurs. In addition, as the sliding speed in the ultrasonic-assisted grinding process caused by ultrasonic vibrations is higher than that in the conventional grinding process, the coefficient of friction reduces [3].

Improvement in coolant supply into the grinding zone is the other reason for the reduction in grinding forces and workpiece temperature when UAG.

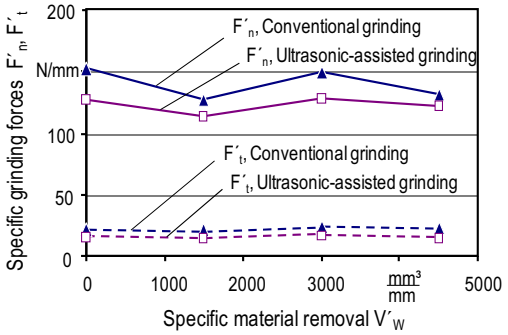


Fig. 5 Specific grinding forces with and without ultrasonic assistance

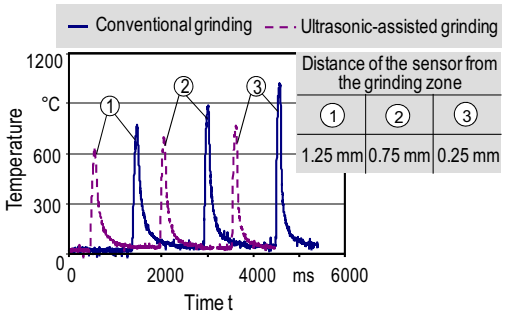


Fig. 6 Grinding temperature with and without ultrasonic assistance

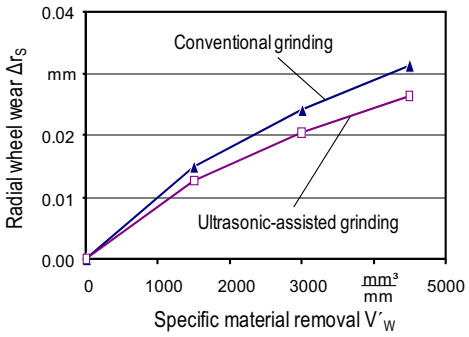


Fig. 7 Reduction in wheel wear due to ultrasonic vibration

The lower grinding forces and workpiece temperature leads to a weaker macro fracture and pullout of grains and consequently to a lower wheel wear when UAG in comparing to CG (Fig. 7).

5. EXPERIMENTAL INVESTIGATION WITH NEW ULTRASONIC DRESSING SPINDLE

The test conditions with new ultrasonic dressing spindle are given in Table 2, where the term "depth of dressing cut" corresponds to the highest theoretical engagement depth of the dressing diamonds in ultrasonic-assisted dressing.

Table 2 Major machining parameters in investigation with ultrasonic dressing spindel

Grinding wheel	Vitrified bond CBN grinding wheel, B151 C150
Workpiece	100Cr6, (60HRC)
Grinding conditions	Feed speed v_{ft} = 3000 mm/min; Cutting speed v_c = 50 m/s; Depth of cut a_e = 0.2 mm
Coolant	5% solution, 100 lit/min
Grinding process	Surface grinding
Dresser	Electroplated diamond form roller D427
Dressing conditions	Overlap ratio U_d = 4; Wheel peripheral speed v_{sd} = 50 m/s; Depth of dressing cut a_{ed} = 2 x 3 and 1 x 6 µm, Dressing speed ratio q_d = ±0.1; ±0.4; ±0.7
Radial ultrasonic vibration of the form roller	Frequency f_{US} =36 kHz, Aplitude A_{US} =6µm

The experimental equipment are the same used in the investigation with new ultrasonic system.

The measured specific normal grinding forces and workpiece roughness values as a function of the depth of dressing cut a_{ed} and dressing speed ratio q_d are shown in Fig. 8. In this figure $A_{US} = 0$ presents the conventional dressing. As can be seen in this figure, ultrasonic assistance causes a reduction in the grinding normal force and an increase in workpiece roughness. This is due to the rougher wheel topography generated by the greater diamond engagement angles and higher volume removed from a grain when impacting with a diamond grain, in ultrasonic-assisted dressing [10]. Performing ultrasonic-assisted dressing in up-cutting mode results in higher grinding forces and lower workpiece roughness values compared with down-cutting mode. This is, in turn, attributed to the flatter diamond engagement paths at a higher relative dressing speed.

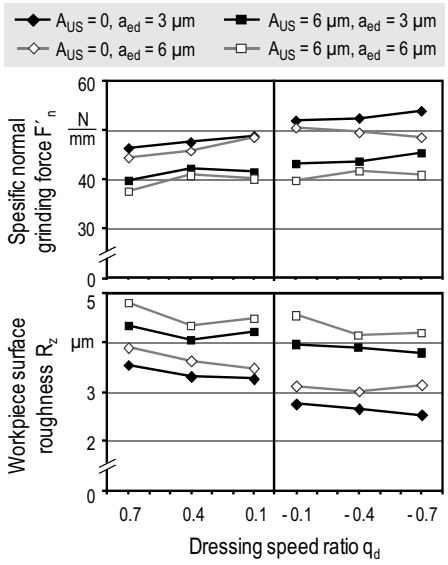


Fig. 8 Reduction in grinding forces and increase in workpiece roughness caused by applying the ultrasonic assistance when dressing with form roller

Fig. 9 shows the grinding wheel wear after conventional and ultrasonic assisted dressing with $q_d = \pm 0.4$ and a specific machining volume of $V'_W = 12000$ mm³/mm. A higher wheel wear after ultrasonic assisted dressing compared to that after the conventional dressing is attributable to the greater engagement angles of the diamonds of dressing roller and lower overlapping of their engagement in ultrasonic assisted dressing. This causes greater crack formations in both CBN grits and bond, which act as the weak points intensifying wheel wear [10]. Interestingly, in Figure 8 the grinding ratios after conventional down-dressing at qd = +0.4 and after ultrasonic-assisted up-dressing at qd = -0.4 are the same.

Fig. 10 shows the radial wear of the roller dresser Δr_R in conventional and ultrasonic-assisted dressing with $q_d = \pm 0.4$ up to a total dressing depth of cut of $a_d = 0.45$ mm. In up-dressing with $q_d = -0.4$, the ultrasonic assistance leads to a reduction in roller wear, and in down-dressing with $q_d = +0.4$ it leads to an increase in roller wear. This is due to the different principal wear mechanisms that are dominant in up-dressing and down-dressing.

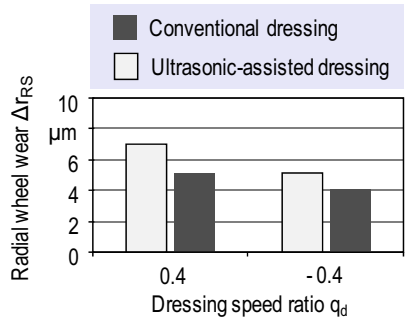


Fig. 9 Radial wheel wear after conventional and ultrasonic-assisted dressing

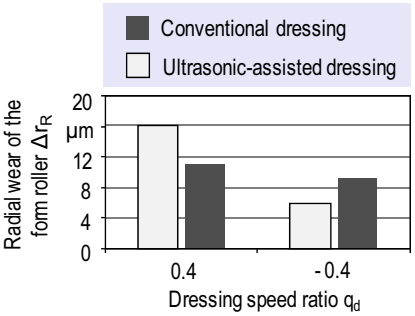


Fig. 10 Radial wear of the form roller with and without ultrasonic assistance

While in down-dressing the microsplinters resulting from the mechanical load are the principal wear mechanisms in the development of wear on the dressing diamonds, in down-dressing the graphitisation of the dressing diamonds arising from the abrasive friction processes is the principal mechanism of wear on the dressing diamonds [Link07, Schu96]. The ultrasonic assistance leads to an increase in the kinematic engagement angle and volume removed from a grain when impacting with a diamond grain, as a result of which the mechanical load on the dressing diamonds increases. On the other hand, the coolant supply into the dressing contact zone is promoted using ultrasonic assistance and therefore, the coefficient of friction between the active dressing partners is reduced. This lowers the development of heat in the dressing contact zone and the intensity of graphitisation. Therefore, in up-dressing, ultrasonic assistance strengthens the principal mechanism of dressing diamond wear, whereas in down-dressing it weakens the wear mechanism [10].

It can be concluded from Figs. 8-10 that in practice, where the conventional dressing of vitrified bond CBN grinding wheels are usually carried out in down-dressing mode (i. e. $q_d > 0$), with ultrasonic assisted up-dressing (i. e. in the range $qd < 0$) a wheel topography comparable to that obtained with conventional down-dressing can be produced; with much lower form roller wear values.

6. CONCLUSION

In this paper, a new ultrasonic system with an innovative block sonotrode was introduced into the ultrasonic assisted grinding. The test results show that the application of the new ultrasonic system can reduce the grinding forces, surface roughness and thermal damages on the workpiece, as well as the wheel wear.

A new ultrasonic dressing spindle was also introduced into the ultrasonic assisted dressing with form rollers. The performed tests show that the ultrasonic assistance can significantly improve the dressing with form rollers in term of roller wear value if the process parameters are set properly.

7. REFERENCES

[1] E. Uhlmann: Surface formation in creep feed grinding of advanced ceramics with and without ultrasonic assistance, Annals of CIRP, Vol.47/1, 1998, pp.249-52.

[2] N. -A. Daus.: Ultraschallunterstütztes Quer-Seiten-Schleifen, PhD Dissertation, Technische Universität Berlin, 2004.

[3] T. Tawakoli and B. Azarhoushang: Influence of ultrasonic vibrations on dry grinding of soft steel, International Journal of Machine Tools and Manufacture, Volume 48, Issue 14, November 2008, pp. 1585-9.

[4] D.E. Brehl and T.A. Dow: Review of vibration-assisted machining, Precision Engineering 32 (2008) 153–172.

[5] Tawakoli, T.; Rasifard, A.: Dressing of Grinding Wheels, in "Machining with Abrasives" edited by M. J. Jackson and J. P. Davim, Springer-Verlag New York, 2011, ISBN 978-1-4419-7301-6.

[6] Y. Ikuse, T. Nonokawa, N. Kawabata, T. Kamo, Y. Yuzawa and K. Unno: Development of new ultrasonic dressing equipment, Int. J. Japan Soc. Prec. Eng., Vol. 30, No. 3 , 1996, pp. 217–222.

[7] M. Nomura, Y. Wu, M. Kato, and T. Kuriyagawa: Effects of ultrasonic vibration in truing and dressing of CBN grinding wheel used for internal grinding of small holes, Key Engineering Materials, Vols. 291–292, 2005, pp. 183–188.

[8] F. Jiao, B. Zhao, X. S. Zhu and and Q. T. Fan, Q.T.: Ultrasonic dressing of grinding wheel and its influence on grinding quality, Key Engineering Materials, Vols. 304–305, 2006, pp. 62–65.

[9] T. Tawakoli, A. Rasifard B. Azarhoushang: Dressing of CBN grinding wheels with ultrasonic assistance, Int. J. Mechatronics and Manufacturing Systems, Vol. 1, No. 4, 2009, pp.321–331.

[10] A. Rasifard: Ultraschallunterstütztes Abrichten von keramisch gebundenen CBN-Schleifscheiben mit Formrollen, PhD Dissertation, Stuttgart University, 2011.

[11] B. Azarhoushang: Intermittent grinding of ceramic matrix composites, PhD Dissertation, Stuttgart University, 2011.

[12] W. Qu, K. Wang, M. H. Miller, Y. Huang and A. Chandra: Using vibration assisted grinding to reduce subsurface damage, Precision Engineering Journal, 24 2000, pp. 329–37.

DESIGN OF AN AERODYNAMIC MEASUREMENT SYSTEM FOR UNMANNED AERIAL VEHICLE AIRFOILS

L. Velázquez-Araque[1] and J. Nožička[1]

[1] Department of Fluid Dynamics and Power Engineering, Faculty of Mechanical Engineering
Czech Technical University in Prague, Czech Republic

ABSTRACT

This paper presents the design and validation of a measurement system for aerodynamic characteristics of unmanned aerial vehicles. An aerodynamic balance was designed in order to measure the lift, drag forces and pitching moment for different airfoils. During the design process, several aspects were analyzed in order to produce an efficient design, for instance the range of changes of the angle of attack with and a small increment and the versatility of being adapted to different type of airfoils, since it is a wire balance it was aligned and calibrated as well. Wind tunnel tests of a two dimensional NACA four digits family airfoil and four different modifications of this airfoil were performed to validate the aerodynamic measurement system. The modification of this airfoil was made in order to create a blowing outlet with the shape of a step on the suction surface. Therefore, four different locations along the cord line for this blowing outlet were analyzed. This analysis involved the aerodynamic performance which meant obtaining lift, drag and pitching moment coefficients curves as a function of the angle of attack experimentally for the situation where the engine of the aerial vehicle is turned off, called the no blowing condition, by means of wind tunnel tests. The experiments were performed in a closed circuit wind tunnel with an open test section. Finally, results of the wind tunnel tests were compared with numerical results obtained by means of computational fluid dynamics as well as with other experimental references and found to be in good agreement.

INTRODUCTION

According to the evolution of unmanned aerial vehicles, commonly referred to as UAVs, several investments have been increasing every year, especially in the field of aerodynamic characteristics which can be obtained through wind tunnel tests.

In the Laboratory of Fluid Mechanics at the Czech Technical University in Prague, particularly in the branch of aerodynamics, a measurement system to get aerodynamic data is needed. Based in the methodology of A. Suhariyomo et al. and J. Barlow et al., this paper presents the design of an aerodynamic characteristic measurement system for UAVs which consists mainly in the design of an aerodynamic wire balance. The balance and the wind tunnel complete the data acquisition system which is verified to assess whether the measurement system is suitable for measuring the aerodynamic performance of different airfoils. The verification was performed with a two dimensional NACA 2415 airfoil, comparing it with reference results in [1,6] and [7], measuring lift, drag forces and pitching moment. In order to go further in the validation of the measuring system, four modified models of this base airfoil were made and tested in the balance; the modification consisted in a step located in the suction surface of each airfoil at 30, 40, 50 and 60 percent of the cord. Lift, drag and pitching moment coefficients versus the angle of attack were obtained for all the models tested and were compared to numerical results obtained in [7]. The whole process is described in the following sections.

DESIGN OF THE BALANCE

This aerodynamic balance was designed for testing the performance of different airfoils for unmanned aerial vehicles measuring lift, drag forces and pitching moment, special attention was paid in this stage because since UAVs fly at low speeds, the magnitude of the forces will be very small, therefore important accuracy is required [5].

Description

The structure of the balance was designed according to the dimensions of the open wind tunnel test section using commercial CAD software; the selected parts for the frame were aluminum profiles due its weight and easy connectivity between the other parts as seen in Figure 1. The balance is composed of six high precision digital hanging scales with a capacity of 50N and a resolution of 0,02N located according to Figure 2.

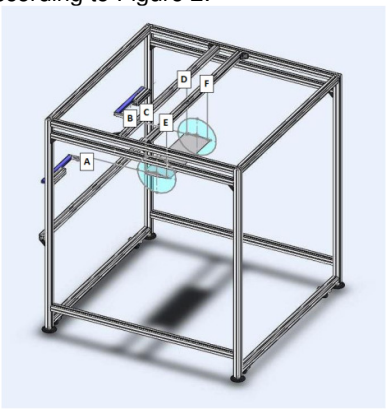


Figure 1. Structure of the aerodynamic balance.

Six forces are measured in scales A, B, C, D, E and F. The wires attached to A and B are parallel to the incoming velocity vector and define a plane which is taken as a reference plane for the balance (x-y plane), these wires point in the x direction.

Figure 2. Location of scales in the aerodynamic balance.

The wires attached to C and D are in a plane that is perpendicular to the x-y plane, which is designate the y–z plane. Wires containing A and C are attached to a common point on the left side of the wing. Wires

containing B and D are attached to a common point on the right side of the wing. Finally wires attached to E and F are parallel to those ones attached to C and D.

The airfoil is attached to two endplates of transparent polycarbonate as shown in Figure 3, these endplates have the main function of reducing considerably the strength of the tip vortices and the induced drag by blocking the leakage around the wing tips [2].

Figure 3. Endplate detail

The balance can produce angle of attack (AOA) changes from 0 to 16 degrees manually with the smallest increment of 2 degrees.

MEASUREMENT SYSTEM APPARATUS

The measurement system includes three components, a low speed wind tunnel system, the aerodynamic balance and data processing software.

Wind Tunnel System

The closed-circuit wind tunnel has an open test section of 750 x 550 mm cross section. It is assembled from straight parts of a closed-return passage with rectangular cross section, elbows with corner-vanes, a rectangular settling chamber and a nozzle. The honeycomb and two screens are placed in a closed-return passage as shown in Figure 4. The maximum air velocity of 17 m/s can be obtained in the test section. A 55 kW three-phase induction motor and a frequency changer are coupled with fan. The wind tunnel has a turbulence intensity at a velocity of 7,5m/s of 1,3% [11]. The flow velocity was measured directly with an anemometer vane type.

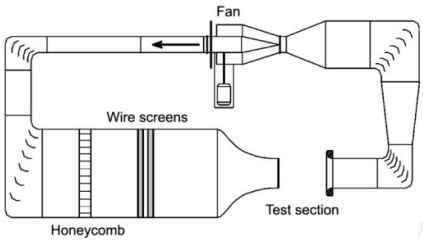


Figure 4. Low speed wind tunnel.

Data acquisition and processing

The procedure for the acquisition and processing the data is in agreement to the procedure for wire balances, attaching the model in an inverted position (upside down) so that aerodynamic lift added to the weight to prevent unloading the wires as the resulting tension can never be allowed to diminish to zero as reviewed in [6].

Since the horizontal wires A and B cannot transmit bending, the vertical force perpendicular to the flow velocity vector, the lift, is obtained from the sum of the forces in the vertical wires:

$$F_L = C + D + E + F \quad (1)$$

The drag is the sum of the forces in the two horizontal wires parallel to the direction of the flow velocity:

$$F_D = A + B \quad (2)$$

The pitching moment is about the *y* axis is given by:

$$M_P = (E + F) \times b \quad (3)$$

where b is the distance between the lines containing C, D and E, F.

All obtained data is collected manually and then processed using calculation software. Then, the general equations to calculate lift, drag and pitching moment coefficients were used as follow:

$$C_D = \frac{F_D}{\frac{1}{2}\rho V^2 A} \quad (4)$$

$$C_L = \frac{F_L}{\frac{1}{2}\rho V^2 A} \quad (5)$$

$$C_M = \frac{M_P}{\frac{1}{2}\rho V^2 AC} \quad (6)$$

Calibration of the measurement system

Since the lift is the largest force by far in a typical aircraft complete model wind tunnel work, extreme care must be taken to ensure that it is orthogonal to the other components [9]. Therefore, the most precise perpendicularity between the wires had to be maintained; otherwise some component of the drag could appear improperly in the lift and vice versa. Concerning the scales, a calibration testing process with low weights from 0,30N to 1,00N was held to observe if each scale was, in fact, measuring what it indicated. (Table 1). The resultant deviation of the scales was ± 0,02N. This possible measurement error for the used conditions represents approximately a change of 0,005 in the computed coefficients.

WIND TUNNEL TESTS

Description

In order to validate the measurement system, low Reynolds number wind tunnel tests on a two dimensional NACA airfoil and four modified models of this one were performed, these results were compared to reference experimental and numerical data.

A NACA 2415 airfoil, which has become increasingly popular on ¼ scale pylon racers [1], was tested as seen in Figure 5. The coordinates were obtained from [4].

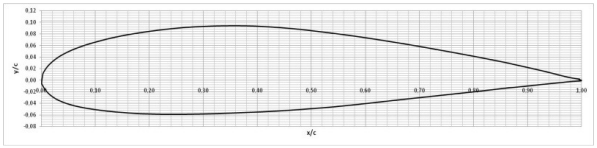


Figure 5. NACA 2415 airfoil

The modification is based mainly on the creation of an abrupt step on the suction side of the original NACA 2415 airfoil. This step simulates a blowing propulsive outlet of the wing in normal flight conditions. The height of the step must be long enough in such a way that does not risk the stability of the structure of the wing model, for that reason the height of the step was defined as one third of the maximum thickness of the original airfoil which means 5 per cent of the chord. It is important to mention that the height of the step was kept constant to focus mainly on the influence of the location of the step along the chord line.

Four different configurations where designed which involved the location of the step at different strategical points chordwise. These points are:

- At the location of the maximum thickness: 30% of the chord (2415-3).
- At the location of the maximum camber: 40% of the chord (2415-4).
- Before the transition point (at 0 AOA): 50% of the chord (2415-5).
- Passed the transition point (at 0 AOA): 60% of the chord (2415-6).

Therefore, four models were built and also the original NACA 2415. Following the NACA numbering system, it was proceeded to assign an additional digit which indicates the position of the step chordwise in tenths of the chord.

The following equations were developed to obtain the ordinates passed the step, since the abscises and ordinates of the original airfoil are given with respect to the origin (0,0),

for $x=x_S$:

$$y_{Um_{x=x_S}} = [y_U]_{x=x_S} - h_S \tag{7}$$

for $x>x_S$:

$$y_{Um_{x>x_S}} = \left[\frac{y_U - h_S}{y_U}\right]_{x=x_S} \cdot [y_U]_{x>x_S} \tag{8}$$

where:

- y_{Um} is the modified ordinate of the upper surface.
- x_s is the abscise of the step.
- h_s is the height of the step.

In Figure 6 it is possible to see all the parameters involved in the modification of the base airfoil, in this case where the step is located at the maximum thickness point.

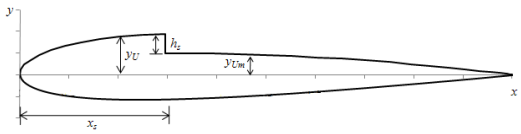


Figure 6. Parameters involved in the modification of the base airfoil

The resulting airfoils after the corresponding modifications are illustrated in Figure 7 a,b,c,d.

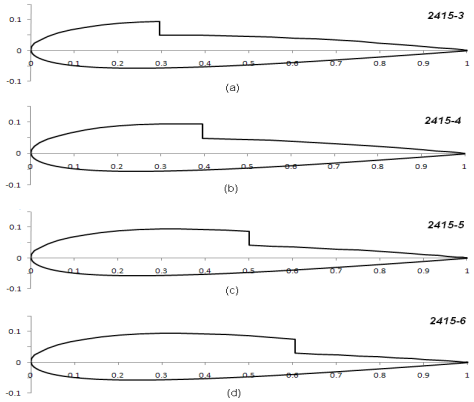


Figure 7. Airfoils developed for testing (a) 2415-3, (b) 2415-4, (c) 2415-5, (c) 2415-6.

The NACA 2415 and all developed airfoil models built for testing (Figure 7 a,b,c,d) have a chord length of 200 mm and a span of 600 mm, which means an aspect ratio of 3. The body of the models was made of expanded polystyrene foam (EPS) and coated with a layer of solid polystyrene as shown in Figure 8 a,b.

(a) 2415-5 (b) NACA 2415

Figure 8 a,b. Airfoil models for testing.

In order to run the tests, the aerodynamic balance was fixed and aligned in the wind tunnel open test section, then the model previously attached to both endplates by means of screws and nuts was hanged in the four scales for lift which were attached to the balance structure and finally it was connected to the two drag scales forming an angle of 90 degrees. The wind tunnel tests were performed at a Reynolds number of 100000 which means a velocity of 7,7 m/s according to the climatic conditions of the measurement days. The angle of attack was set from 0 to 16 degrees with an increment of 2 degrees [8]. The experimental results are validated with numerical results which will be also described later in the following sections.

Adjustments

In order to obtain the aerodynamic forces applied only on the model, the drag of the endplates and the rod which hold the model in the model had to be obtained because it is also being measured by the drag scales, for this purpose the endplates attached

with the rod where mounted and the drag force was measured by scales A and B.

Then, the drag force of the rod was computed by means of Eq. (4) using as its drag coefficient the one obtained in reference with the corresponding Reynolds number.

RESULTS AND DISCUSSION

In the following graphs (Figures 9 and 10) the aerodynamic performance of a model of a NACA 2415 airfoil is shown, experimental and numerical data from references [1,6] and [7] respectively are also included.

In the lift coefficient graph shown in Figure 9, it can be observed that the stall region appears for an angle of attack of 16 degrees with a C_L of approximately 0,96. In this point, the measured C_L is higher that the C_L from experimental reference in [6] at Re=340000. However it is slightly lower than the reviewed in the numerical reference in [7] at Re=100000. Comparing to reference in [1] at Re=100000, all values of the measured C_L are considerably lower.

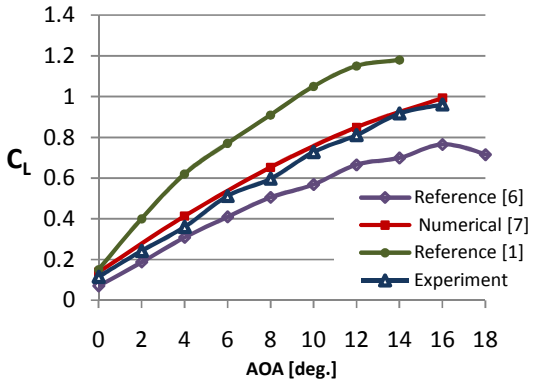


Figure 9. Lift coefficient graph of NACA 2415 airfoil.

In Figure 10 it is possible to see the drag coefficient as a function of the AOA for the NACA 2415 airfoil, the measured C_D is in very good agreement with numerical reference. Comparing to the C_D reviewed in [6], all the values are very similar but these latter ones are slightly smaller, from an AOA of 16 degrees the C_D increases considerably but this value could not be compared because as mentioned before, the highest possible AOA is 16 degrees so far. Concerning the reviewed reference in [1], all the values are considerably smaller than the measured ones and all other references. These discrepancies in drag are assumed to be primarily caused by imperfections in the airfoil construction and the inaccuracies in the experiment.

Figures 11 - 14 show C_L versus AOA, C_D versus AOA, C_L versus C_D and C_M versus AOA of four wings which have a NACA 2415 airfoil as a base shape and a step on the suction side located according to Figure 7. Each graph also presents obtained experimental results of the original NACA 2415 for further comparison. All the wings were tested at a Reynolds number of 100000.

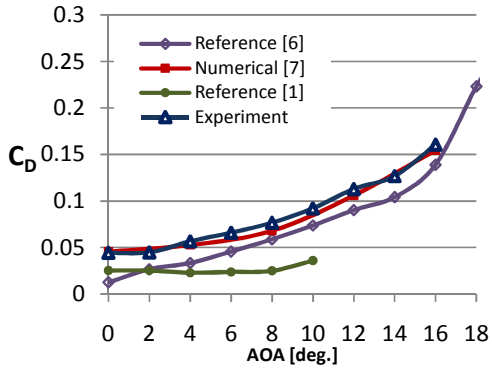


Figure 10. Drag coefficient graph of the NACA 2415 airfoil.

Concerning the lift coefficient, shown in Figure 11, it can be seen that the location of the step does not have influence on the slope of the lift curve, however as the step moves towards the trailing edge, the lift curve moves upwards, therefore this affects the values of the lift coefficient. The 2415-3 airfoil has the lowest lift for all angles of attack and its behavior is close to a symmetric airfoil (almost zero lift at zero angle of attack). The curve is linear until 14 degrees of AOA where an abrupt decrease begins which corresponds to the stall point.

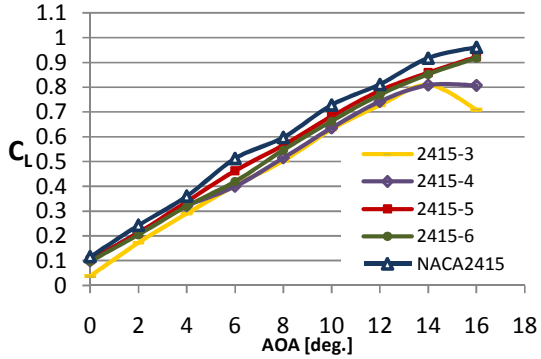


Figure 11. Experimental lift coefficient graph for all airfoils models.

The 2415-4 airfoil lift coefficient at zero angle of attack is approximately higher by 0,07 than the corresponding value for the 2415-3. The lift curve of the airfoil is not linear, at 6 degrees of AOA the curve follows the same path of the one for the 2415-3 airfoil curve until an AOA of 14 degrees where a low decrease starts. The results show that the 2415-3 and 2415-4 airfoils have approximately the same stall point at 14 degrees of AOA but after this point, the decrease in lift for 2415-3 airfoil is much more abrupt.

The lift coefficient behavior for 2415-5 and 2415-6 airfoils is very similar. The curves start at the same value of C_L, like 2415-4 airfoil but their path is approximately linear until 12 degrees of AOA where the slope begins to shrink moderately. The presence of the stall point for these curves is not shown very clear but it can be assumed that is very next to 16 degrees of AOA. The original airfoil lift curve has slightly higher values of C_L for all AOA but its behavior is almost the same; the

stall region begins after 15 degrees of AOA. These characteristics were expected since the step produces a sudden change in the pressure distribution along the upper side of the airfoils which contributes to decrease the lift.

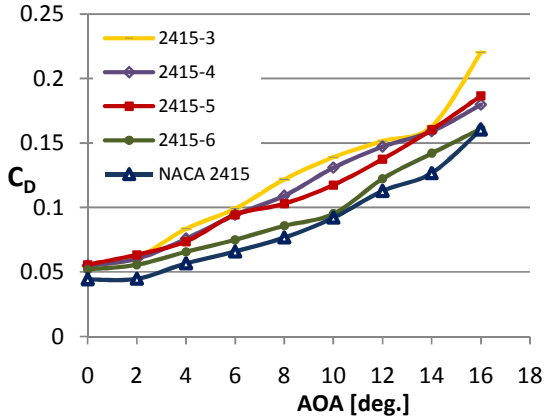


Figure 12. Experimental drag coefficient graph for all airfoils models.

In Figure 12 above it is possible to see the drag coefficient versus the AOA for all the models tested. All of the modified airfoils reached approximately the same value of drag coefficient at zero angle of attack. The drag coefficient of the original airfoil value is about 0,01 lower. A similar behavior of all drag curves is seen at 2 degrees of AOA.

The 2415-3, 2415-4 and 2415-5 airfoils presented similar drag characteristics for low AOA, specifically until 6 degrees, after this angle the values of C_D are different until 14 degrees of AOA where the three curves present almost the same value of C_D. Also, between 6 and 14 degrees of AOA the 2415-3 airfoil presents high values of C_D followed by the 2415-4 and 2415-5. After 14 degrees of AOA, the drag for 2415-4 and 2415-5 is similar and the drag of 2415-3 airfoil showed a substantial increment.

The 2415-6 airfoil drag curve is similar to the original NACA 2415 airfoil curve; there is only certain shift at 10 degrees of AOA where both drag coefficient values are almost the same. From this AOA the drag of 2415-6 increases again until 16 degrees of AOA where the C_D value of both airfoils equals. It is clear that the 2415-6 airfoil presented the best drag performance among all modified airfoils tested.

The polar diagram shows a lot of information in a very compact format; therefore it is always considered the most important part of the results [10]. In Figure 13 it is possible to see the polar curves of all airfoils tested. One of the important points in this diagram is the optimum glide ratio (OGR) which is obtained drawing a line from the origin (C_L=0, C_D=0) tangentially to each curve according to the diagram the results are:

- NACA 2415: OGR=7,778 at 9 degrees of AOA.
- 2415-6: OGR=6,947 at 10 degrees of AOA.
- 2415-5: OGR=6,0 at 10 degrees of AOA.
- 2415-4: OGR=5,161 at 13 degrees of AOA.
- 2415-3: OGR=5,142 at 13 degrees of AOA.

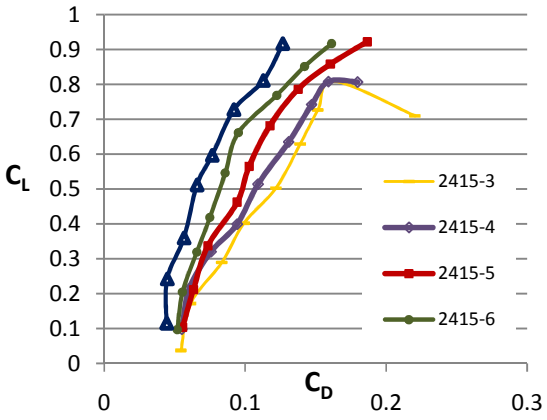


Figure 13. Experimental polar graph for all airfoils models tested.

According to this the curves can be divided in two groups. In the first group (2415-3 and 2415-4 airfoils) the stall points were achieved at 14 of AOA and the curves are convex between 4 and 14 degrees of AOA. In the second group (2415-5 and 2415-6 airfoils) the curves are convex in a very small range between 4 and 8 degrees of AOA and the stall point was not achieved at the measurable scale. The original airfoil polar is concave in almost the whole range of AOA. This diagram also can show a general idea of the aerodynamic performance of airfoils and it is possible to see that while the location of the step chordwise moves to the trailing edge, the aerodynamic performance of the airfoil tested increases.

The pitching moment coefficient was obtained by means of Eq. (6) and it is computed with respect to the leading edge according to the aerodynamic balance as shown in Figure 14. In this chart it is possible to observe that the differences between the NACA 2415, 2415-6 and 2415-5 airfoils are minimal. The 2415-4 airfoil presents a lower moment until 4 degrees of AOA and after this point the moment values are similar to 2415-5, 2415-6 and NACA 2415 airfoils. The 2415-3 airfoil presented a lower pitching moment along all AOA.

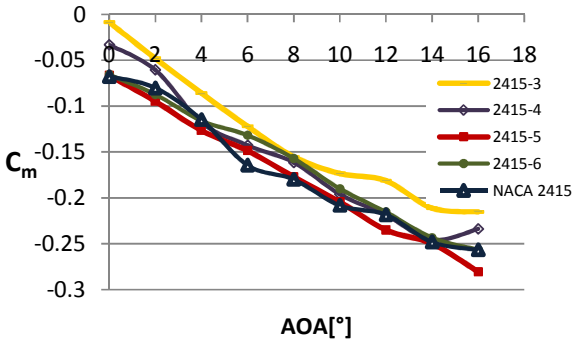


Figure 14. Experimental pitching moment coefficient at the leading edge for all airfoils models tested.

In Figure 15 it is possible to see the pitching moment coefficient with respect to a point located at 25% of the chord from the leading edge, this point is usually approximated as the aerodynamic center. This

moment was computed by means of the following equation:

$$(C_m)_{1/4} = C_m + \frac{1}{4}C_L \quad (9)$$

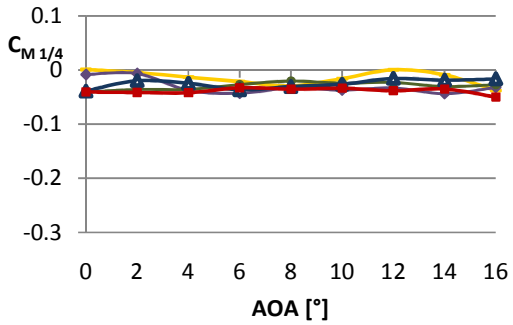


Figure 15. Experimental pitching moment coefficient at 25% of the chord for all airfoils models tested.

In theory, the aerodynamic center is the point at which the pitching moment coefficient for the airfoil does not vary with the angle of attack [10]. It can be seen also in Figure 15 that the pitching moment coefficient for all models tested had very small differences with respect to the AOA, it was negative which indicates a pitch down tendency and the values were in a range between 0,00 and -0,04.

In Figure 16 it is possible to observe the lift to drag ratio C_L/C_D versus AOA, this diagram is used to see the performance of all airfoils tested and along all AOA, in this manner it is clear that all modified profiles reached their maximum C_L/C_D in a smaller range angle of attack comparing to the NACA 2415 airfoil, this one presents a maximum C_L/C_D of 8 for a range between 6 and 10 degrees of AOA.

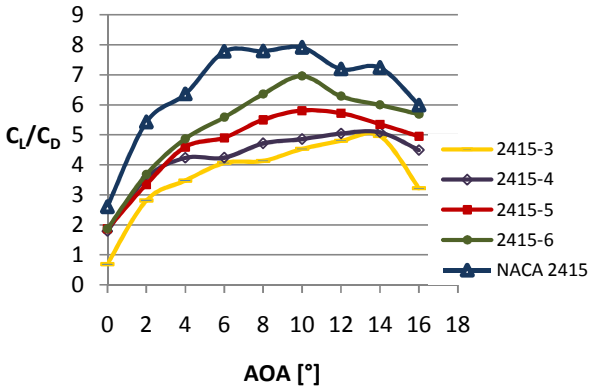


Figure16. C_L/C_D versus AOA for all airfoil models tested.

The maximum C_L/C_D for 2415-6 and 2415-5 airfoils is 7 and 6 respectively at 10 degrees of AOA. The 2415-4 and 2415-3 airfoils presented a maximum C_L/C_D of approximately 5 in a range between 10 and 14 degrees of AOA.

CONCLUSION

A measurement system for aerodynamic characteristics of unmanned aerial vehicles was designed including a low speed wind tunnel and the aerodynamic balance reviewing previous investigations. The balance was designed and built in this study. In order to validate the design a wing model of a NACA 2415 airfoil was built and tested, the results were compared to experimental and numerical reference data and found in good agreement. Then four modified airfoils were built with and tested, the results showed that the aerodynamic performance of the airfoil increases when the step moves towards the trailing edge. As further study, it is recommended to design a system which allows performing tests of these airfoils with blowing and make comparisons with results obtained in this paper. To sum up, according to the results in this research, it can be concluded that the aerodynamic measurement system is acceptable for measuring the performance for different wing models for unmanned aerial vehicles.

REFERENCES

[1] Selig M.S., Lyon C.A., Giguere P., Ninham C., Guglielmo J.J.(1996): "Summary of Low-Speed Airfoil Data", Vol. 2, Viginia Beach, SoarTech Publications.

[2] Barlow J.B., Rae W. H., Pope A.(1999): "Low-Speed Wind Tunnel Testing", 3rd ed., New York, John Wiley & Sons, Inc.

[3] Suhariyono A., Hyun Kim J., Seo Goo N, Cheol Park H., Joon Yoon K.(2005): "Design of Precision Balance and Aerodynamic Characteristic Measurement System for Micro Aerial Vehicles", Konkuk University.

[4] Abbott I.H., Von Doenhoff A.E.(1959): "Theory of Wing Sections", New York, Dover Publications Inc.

[5] Pope A.(1954): "Wind Tunnel Testing", 2nd ed., New York, John Wiley & Sons, Inc.

[6] Ghods M.(2001): "Theory of Wings and Wind Tunnel Testing of a NACA 2415 Airfoil", Technical Communication for Engineers, University of British Columbia.

[7] Velazquez L. (2010): "Development of an airfoil for an unmanned aerial vehicle with internal blowing propulsion system".PhD Thesis. CTU in Prague.

[8] Velazquez L., Nožička J.(2009): "Low Speed Wind-Tunnel Testing for an Unmanned Aerial Vehicle", Proceedings of Colloquium of Fluid Dynamics 2009, Prague, 2009.

[9] Tropea, Yarin, Foss (Eds.)(2007): "Springer Handbook of Experimental Fluid Mechanics", Springer.

[10] Clancy L.J.(1986): "Aerodynamics", Sterling Book House, Mumbai.

[11] http://profily.fs.cvut.cz/, Laboratory of Department of Fluid Dynamics and Power Engineering of the CTU. Section Airfoils and Straight Blade Cascades.

Large-area Overhead Manipulator for Access of Fields

Jeffrey W. White, Roger V. Bostelman

***Abstract*—Multi-axis, cable-driven manipulators have evolved over many years providing large area suspended platform access, programmability, relatively rigid and flexibly-positioned platform control and full six degree of freedom (DOF) manipulation of sensors and tools. We describe innovations for a new six DOF manipulator, called 'Large-area Overhead Manipulator for Access of Fields (LOMAF)' that is highly scalable and provides access to areas that can measure from several square centimeters to several hectares and can traverse level or irregular terrain, including vertical surfaces. Two scaled prototype have been developed and tested providing the basis for designing a much larger, computer-controlled system. The LOMAF design, prototypes, tests, results and suggested applications are presented in this paper.**

I. Introduction

A. Current Field Access Methods and Issues

IN many types of field research, including agronomy, natural resource management, environmental science, and ecology, situations arise where minimally invasive measurements are required over field areas on the order of 10,000 m^2 (107600 ft^2). Typically, contact with the soil surface must be minimized, precluding use of wheeled vehicles or frequent entry by personnel on foot. Diverse types of sensors and imaging tools can perform the requisite measurements, but a means to position the instruments is needed. Current options include fixed and rotary-winged aircraft, aerostats and cranes. While each has niche applications, they all have limitations related to cost of operation, inability to rapidly position sensors near the field surface and to operate at night or under inclement weather.

The fundamental challenge thus is how to position an instrument platform over a potentially irregular surface, providing positional accuracy of 50 mm (2 in) or less, using a system that has low operating costs and can function continuously. Cable-suspended manipulators provide an attractive alternative, but current designs require support towers whose heights are approximately half the longest axis of the target area. Furthermore, these manipulators have a relatively large "dead" area near each of the support posts which the manipulator cannot access without increasing altitude or tilting the instrument platform.

Manuscript received March 10, 2010. This work was supported in part by the U.S. Department of Agriculture, Agricultural Research Service and the U.S. Department of Commerce.

J.W. White. is with the Department of Agriculture, Agricultural Research Service, Arid Land Agricultural Research Center, 21881 N Cardon Lane, Maricopa, AZ 85138, USA. (phone: 520-316-6368; email: Jeffrey.White@ars.usda.gov).

Roger Bostelman is with the Department of Commerce, National Institute of Standards and Technology, Intelligent Systems Division, Gaithersburg, MD 20899 USA (phone: 301-975-3426; e-mail: roger.bostelman@nist.gov).

B. RoboCrane

The National Institute of Standards and Technology (NIST), Manufacturing Engineering Laboratory (MEL) Intelligent Systems Division (ISD) developed a cable-suspended manipulator called RoboCrane [1] over many years and applied it to a variety of applications [2]. The RoboCrane is a cable driven, multi-purpose manipulator based on the Stewart Platform Parallel Link Manipulator. It provides six degree-of-freedom (DOF) load control via teleoperation, graphic off-line programming, hybrid control modes and computer programs for fully autonomous maneuverability [3]. The RoboCrane was first developed under a Defense Advanced Research Project Agency (DARPA) contract to stabilize loads suspended from conventional cranes. Current configurations have advanced to include land, sea, air-lifted, and space applications. It can be designed for high lift-to-weight ratio, stable gantry or other superstructure (e.g., masts) configurations, flexibility, precise maneuverability, and mobility over a variety of surfaces including very rough terrain.

The RoboCrane consists of a stable platform supported by six cables suspended from three base points on a fixed or mobile structure. The six cables are arranged to kinematically constrain the stable platform such that its stiffness is determined by the tensile elasticity of the cables. Maximum stiffness is maintained so long as perturbing forces and/or torques are below a threshold determined by the weight of the load. For forces or torques above that threshold, one or more cables will go slack, and stiffness will drop to that generated by pendulum forces of the load on the cables remaining taut. When all six cables are in tension, the stable platform is kinematically constrained, and there exists a known mathematical relationship between the lengths of the six cables and the position and orientation of the platform. The lengths of the six cables are controlled by six winches. These are controlled and coordinated by a computer. Input commands can be from a joystick or from a computer programmed to maneuver the stable platform and onboard tools or sensors over a very large work volume. The RoboCrane platform centroid work volume with only gravity loading is generally defined by an inscribed circle within an imaginary triangle joining the upper three suspension points as shown in Figure 1. A suspended, uniformly distributed load platform from three points can access approximately 60% of the triangular area formed by the base points The inscribed square region to be accessed has about 36% of the area of the bounding triangle.

This paper will describe a case application from plant sciences followed by a description of two different LOMAF designs - with and without downhaul cables. Experiments that compare the two designs and their results are then explained.

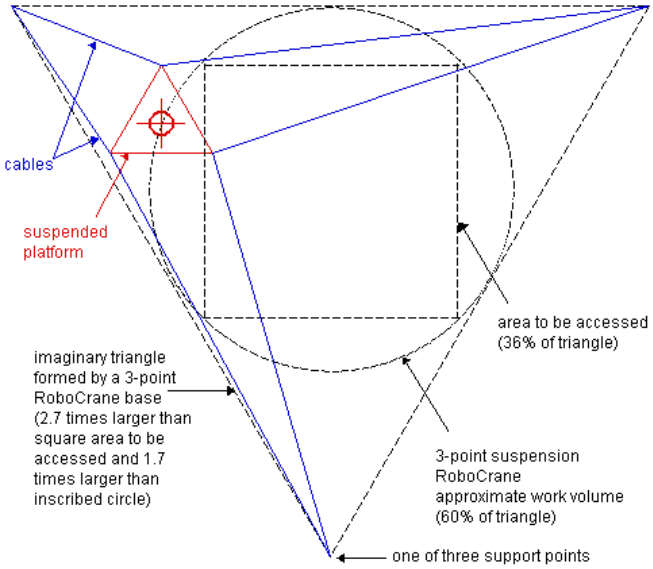


Figure 1 – Approximate accessible areas (circle and square) of a 3-point base (triangle) RoboCrane suspended platform's center of gravity.

II. Plant Science Case Study and LOMAF Hypothesis

A. A Case Application from Plant Sciences

The grand challenge of plant science and crop improvement in the 21st century is to predict plant traits (phenotypes) such as yield and time of maturity from data on genotypes, environment and management. Gene detection and sequencing no longer constrains research in plant genetics. "Next generation DNA sequencing" is expected to allow sequencing of entire genomes at costs of less than US$2 per megabase [4], and high-throughput tools for characterizing germplasm with single nucleotide polymorphisms are attaining prices under US$0.50 per data point. Data on environmental conditions and field management, including experimental treatments, are routinely available. However, rapid and accurate measurement of phenotypes remains challenging, especially for phenotypes that cannot be directly assessed using visual criteria and for traits related to abiotic stresses that require precise control of stress regimes. Associating phenotypes with genotypes for dissection of complex traits and for efficient selection of stable, high yielding varieties requires analyses of hundreds to thousands of genetically distinct plants grown in environments that allow for differential expression of multiple genes.

This situation has led to proposals that "high-throughput phenotyping" (HTP) capabilities be developed for accurately characterizing large numbers of genetic lines or individual plants with a fraction of the time and labor currently required [5]. Accomplishing HTP in a cost-effective manner, however, requires new techniques and infrastructure. The most promising HTP approaches involve using a vehicle to deploy a platform bearing multiple sensors that can quantify plant traits on a time scale of a few seconds per plot. Previous experience in phenotyping for water and nutrient management, and to a lesser extent for breeding and genetics, has established that numerous plant traits can be measured reliably using remote sensing [6].

B. Hypothesis

A suspended platform with onboard sensors and plant care/monitoring equipment is feasible using a RoboCrane design as demonstrated at NIST on a variety of applications. However, obtaining the desired higher ratio of access area to base size requires a new cable configuration. System costs are driven higher with added control axes and therefore the design should include a minimal number of cables and winches, but these costs are partially offset by the reduced mast heights and expanded work volume. The system would require manipulator access near the suspension points, as opposed to RoboCrane's three point suspension geometry. A smaller access area to base size is expected with only gravity loading, as well as minimal platform stability unless additional platform weight is provided above the onboard sensor and equipment platform loading. Our hypothesis is that preloading the platform with downhaul cables should provide a higher access area to base size and require minimal platform loading, but added cables will increase system cost.

III. LOMAF Designs

An innovative six DOF manipulator was developed at NIST and the Arid Land Agricultural Research Center (ALARC), called the Large-area Overhead Manipulator for Access of Fields (LOMAF). The manipulator maximizes the base to accessible area ratio while minimizing the height of supporting masts. LOMAF is highly scalable, providing access to areas that can measure from several square centimeters to several hectares and can operate on level or irregular terrain, including vertical surfaces. Convex and/or concave land surfaces can be autonomously accessed from above after setup without disturbing any surface features throughout the target area. Based on RoboCrane research, control modes could include manual, velocity and force with desired accuracy of within 25 mm (1 in). The manipulator includes four or five cables (without downhaul cables) or eight or nine cables (including downhaul cables) that are independently controlled by hand, weights or powered winches through the use of an operator controlled joystick, graphical user interface or by a computer. Two models have been built to study how a LOMAF is designed to access an 86 m x 86 m (2 acre) crop field, including a 1:160 scale model and a 1:17 scale model. Three specific design cases using these models are considered:

Case 1 - Suspended platform without downhaul cables or running lines

Case 2 - Suspended platform with platform downhaul cables moving along running lines

Case 3 - Suspended platform with platform downhaul cables fixed to the masts.

For a full scale LOMAF, the design is expected to include: four 20 m (65 ft) masts with 6 mm (0.25 in) diameter, 6 x 19 extra improved plow steel wire rope cables with a safe working load of 618 kg (1360 lb) [7]; hinged pulleys mounted to the masts and platform; a 6 m (20 ft) square platform; and electric powered winches with rotary encoders mechanically attached to the winch motors and electrically connected through power amplifiers to a control computer similar to a RoboCrane. Additional tension sensors and/or motor current measurement can be used to provide cable tension or force control, as well as safety overrides. In all LOMAF designs, winches can be mounted to the platform or preferably, to the masts in order to minimize platform mass.

RoboCrane mast or suspension point height is generally set to allow a platform cable angle of 10º or more with respect to the horizontal axis depending upon the application. Below 10º the cable tensions exponentially increase with increased platform lift. The LOMAF workvolume is similar to the top portion of the RoboCrane workvolume having cables at up to 10º with the horizontal and with a maximum platform loading of no more than 50 kg (110 lb). For relatively short mast heights as compared to mast separation distances (MSD), for example a 20 m (65 ft) mast height to 86 m (280 ft) MSD, the platform could access the lower 60% mast height field volume of 12 m (39 ft) high. An approximate Case 1 mast height and cable tension estimate is shown in Figure 2. The LOMAF can also access well below the suspension points similar to RoboCrane, for example in open pit applications, with much higher cable angles and lower cable tensions.

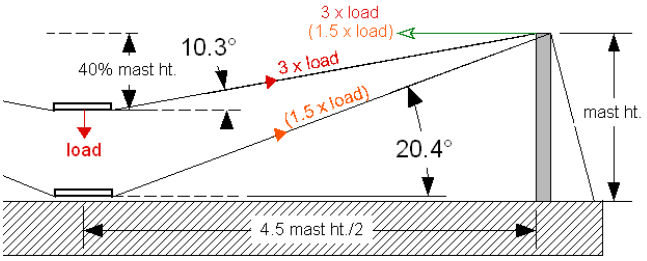


Figure 2 – Partial LOMAF side view of approximate worst case (along a field edge) cable tensions and mast top horizontal loading with respect to cable angle and suspended platform load.

Example lift cable tensions for Case 1 with the platform at the maximum height at the field center are (¼ load)/sin (10º) where 'load' is the total platform weight and the angle is with respect to horizontal. This approximates the cable tension to be (1.5) x full load and nearly equals the lateral mast top loading of (cable tension) x cos (10º). Positioning the platform along an edge causes cable tensions and lateral mast loading of (1/2) load/sin (10º) or nearly (3) x full load as shown in Figure 2. Therefore, a suspended load of 50 kg (110 lb) would produce maximum cable tensions along a field edge of approximately 71 N (317 lb-f) on one of two cables with 70 N (312 lb-f) horizontal mast top loading on one of two masts. Minimal Case 1 lift cable tensions and mast loading are approximately load/sin (90º) with the platform next to a mast. At this platform position, the remaining two of the four total masts experience much less loading. Minimal Case 1 lift cable tensions and mast top loading occur with the platform next to a mast are approximately equal to the load.

Case 3 includes cable tensions resulting from both lift cables and downhaul cables which are controlled to maintain platform level and increase field access. This design's variables of force and cable angles include lift and downhaul cable tensions, lift and downhaul cable attachment heights, and platform weight. These parameters are useful to size masts, cables, winches, and the platform.

A. 1:160th scale model LOMAF

A 1:160 table-top sized, scale model LOMAF without downhaul cables was built at NIST and is shown in Figure 3. The physical working model uses spring-loaded cable spools (black circles) and allows the operator to move the platform manually to test accessible areas of an 86 m x 86 m field. Red dotted cables have been added to Figure 3 to show how two platform yaw control cables would replace one cable. The same yaw control cable design can be used with four downhaul cables totaling nine cables. A nine cable LOMAF design with downhaul cables is shown in Figure 4 a and an eight cable design is shown in Figure 4b as in Case 2.

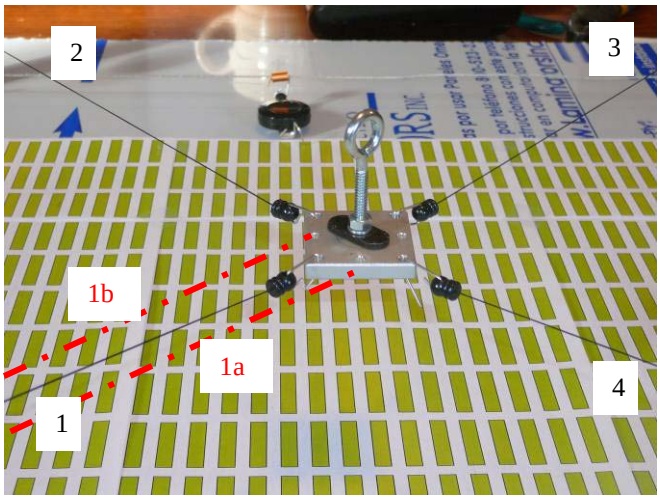
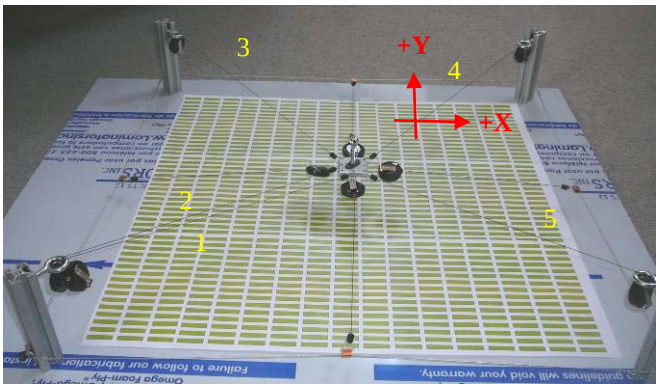


Figure 3 – 1:160 scale model LOMAF without downhaul cables. Cables are numbered 1 through 4. Red dotted cables depict two cables 1a and 1b that can replace cable 1 to control platform yaw rotation.

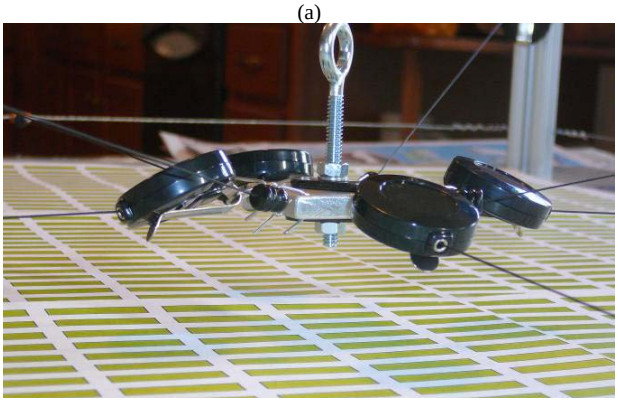
The green rectangles below the platform signify crop beds where a 6 m (20 ft) square platform can reach across two crop beds simultaneously. Running lines, in this design, are connected between each mast and a sliding tube that slides along the running line is connected to downhaul cables connected to the platform. In a full scale LOMAF, a set of rollers would roll along each running line in place of the sliding tube to reduce friction. These cables are expected to be force-controlled while the mast lift cables are expected to be position-controlled.

Without downhaul cables, the spring-loaded model cables appeared to lift a platform corner closest to the nearest mast.

By adding downhaul cables, they pull the platform close to the running line height thereby allowing positive control along the vertical axis. Using downhaul cables, the model platform allows access to all areas of the square 'field' of crop beds including access to each mast, along all edges and throughout the center of the field. For example, as shown in Figure 4 a, if a +X axis downhaul cable and a +Y axis downhaul cable were each given a force command to spool in cable and the lift cables are commanded to position the platform against post 3, the platform can in fact contact the mast. Either an X or Y only force command allows edge access. We therefore, prove that the LOMAF design allows a maximized access area to base size (tower separation).



(a)



(b)

Figure 4 – 1:160 scale model LOMAF with downhaul and running line cables. The black circles are spring-loaded cable spools. The center square is the suspended platform, masts are at the corners and the small rectangles below the platform signify crop beds. The (a) end view and (b) side view showing a platform elevated above crop beds.

Figure 5 shows the table top model tilted to a variety of angles displaying that the LOMAF would be useful on level or sloped areas. In fact, with positive platform control along all axes as in the case of a LOMAF with running lines and downhaul cables, even vertical surfaces can be accessed in example applications such as building construction.

B. 1:17 scale model LOMAF

A 1:17 scale model of an 86 m x 86 m (2 acres) was built at the ALARC to test potential field access and cable tensions. Figure 6 shows the prototype set up as in Case 1.

Figure 5 – Variable angles of the LOMAF showing application to sloped (left) and even vertical (right) surface access.

Figure 6 - Manual 5 m prototype LOMAF at the ALARC.

Masts included four 2.4 m (8 ft) high by 50 mm (2 in) diameter posts with pulleys attached to their tops. Each of four lift cables was attached to a 0.3 m (1 ft) plywood platform corner between the top of each mast and the wooden platform corners. Buckets of weights were tied to the end of each rope to allow the platform to lift and settle in a raised field position. Cables can then be additionally loaded by an operator or by adding bucket weights to move the platform towards a mast. Synthetic polymer ropes were used in place of wire rope cables. Lift cables were made of 6 mm (1/4 in) nylon rope and 4.8 mm (3/16 in) braided polypropylene rope was used for the downhaul cable. Case 3 was also setup on the model using a single downhaul cable attached to one mast base and the platform corner as shown in Figure 7.

IV. Experimentation of Scale Prototypes

Basic issues related to platform leveling, cable tensions and usable work area were tested using the two models. Case 2 was tested using the 1:160 scale LOMAF model by moving the platform by hand to all areas of the field. The critical areas for the platform to reach are near the masts and at the outer field perimeter. By moving the platform to these areas, it was apparent that it could reach all field areas, even the perimeter. Therefore, Case 2 was not tested with the 1:17 scale prototype model. However, Case 2 also requires that 9 cables and winches plus running lines be used to provide this ideal case.

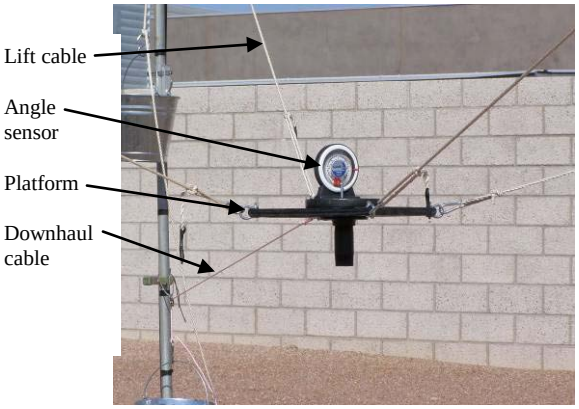


Figure7 - Manual 5 m (20 ft) prototype LOMAF with added downhaul cable attached to a mast.

The 1:17 scale LOMAF was tested with Cases 1 and 3 cable rigging configurations using two different platform weights (1.7 kg (3.7 lb) and 4.0 kg (9 lb)) and maintaining the platform at a constant height of 1.5 m (59 in) below the mast tops. Case 1 experiments included 4 ropes for platform lift and positioning between the platform and each mast top. Case 3 experiments added a downhaul cable between the platform and one mast base with the downhaul attached 0.5 m (19.7 in) below the target platform height (i.e., 2 m below the mast tops). To simplify testing of Case 3, a single downhaul was used to show access and cable tensions near a mast, as shown in Figure 7, and measurements focused on the area from the center of the LOMAF to the post with the downhaul.

The basic experimental procedure consisted of guiding the platform to an approximate X-Y position, while maintaining a constant elevation (Z), and then adjusting weights until a stable position was attained and if possible, the platform was level. Static cable tensions were measured by weighing the buckets when the platform was in a static, equilibrated position. Inclination of the platform was measured with an inclinometer with a resolution of 1° and distances were measured with a tape measure.

V. Experimental Results

The Case 1 1:17 LOMAF relies only on gravity as a vertical force. When the platform was moved from the center position, it pitched and rolled up towards the nearest post, as shown in Figure 6. The platform was shown to lift towards the mast cable pulley as the platform approached it, therefore limiting the work volume for this case. Figure 8 shows a graph of platform inclination above horizontal versus platform distance from the mast with the platform positioned a constant height of 1.5 m (59 in) below the top of the mast.

Figure 7 shows the prototype with both a lift and downhaul cable (Case 3) on one mast and the platform shifted near the mast. Note that in this design, the platform can maintain level even when near a mast. Directly between two masts neither Case 1 or 3 design would provide access along the edge since there is no lateral force pulling the platform towards the edge except for a minimal mast lateral component from lift and downhaul cables in that direction.

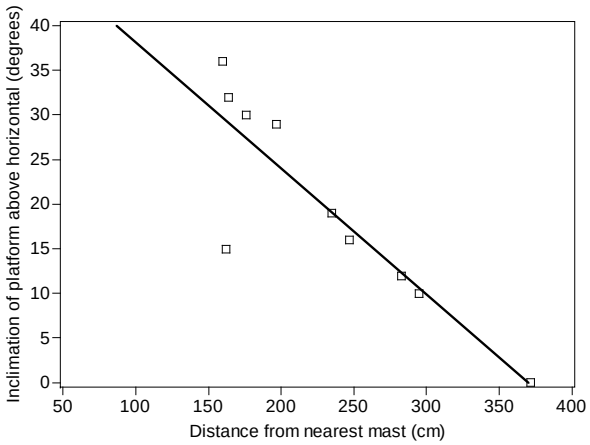


Figure 8 - Graph of platform angle versus distance of the platform to the mast.

This force is zero at the field edge inline with two masts and the platform will not reach the edge while the cable tension rises exponentially. Therefore, a smaller platform work volume is inherent in Case 1. However, Case 3 provides platform access to the masts between the downhaul cable and lift cable attachment heights. Figure 9 shows a top view of approximated work volumes of Case 1 and Case 3.

Figure 10 a and b show plots of cable tensions versus distance from the nearest post with 1.7 kg and 4.0 kg platform weights. For these tests, the platform was maintained level (±1°) using downhauls as in Case 3.

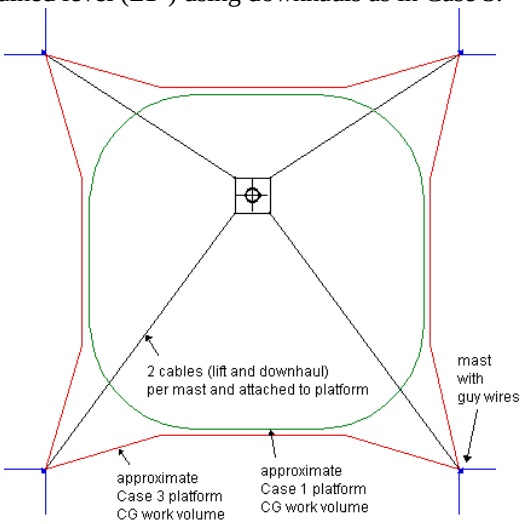
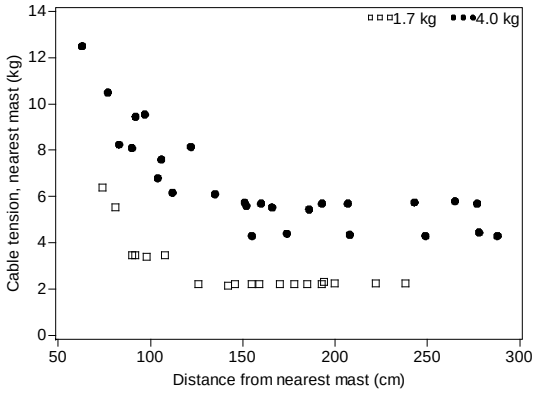


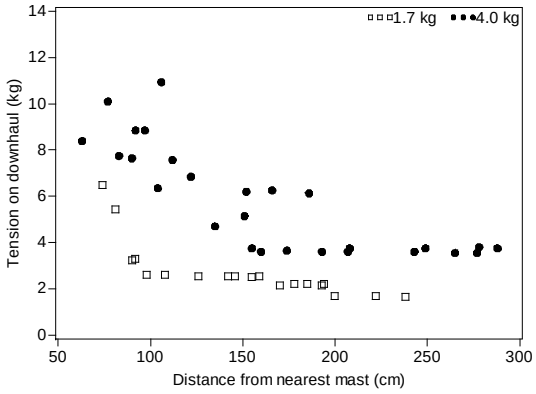
Figure 9 – Approximated work volume comparison of Case 1 and Case 3 LOMAFs.

With a 1.7 kg (3.7 lb) mass, the tension on the downhaul and lift cables remained stable and low up to about 100 cm (39 in) from the post, while with the 4.0 kg (8.8 lb) mass, the tension began to increase at about 150 cm (59 in) from the mast. Assuming for this prototype that the ideal work area is 25 m^2 (2690 ft^2), a reduction in area corresponding to a 100 cm (39 in) diagonal from each mast would reduce the work area to 13 m^2 (140 ft^2), and 150 cm (59 in) diagonal, to 8 m^2 (86 ft^2). The results thus confirm that the downhaul line provides an effective method to level the platform, with the

important caveat that as the platform load is increased, the effective work area will be reduced (or that the system will have to be engineered for tensions considerably greater than the equivalent mass of the platform). Also note that the tensions decrease as the platform moves away from the mast indicating that the load is shared by the other lift cables.



a



b

Figure 10 - Comparison of tensions on cables using a platform mass of 1.7 kg and 4.0 kg for: a. Cable to nearest mast, b. Downhaul, versus distance from the nearest mast.

Regardless of the platform loading, downward forces caused by downhaul cables were determined to control platform level as needed to maintain similar sensor-to-crops distances throughout the entire work volume.

VI. Additional Applications

The LOMAF has proven to be a viable field access system for level or sloped fields. Below are a list of potential LOMAF (underlined) applications and their associated potential tasks:

Agricultural production and research: Crop Monitoring, Seeding, Spraying, Harvesting, Pruning
Mine Detection: Scanning, Finding, Detonating, Accessing, Monitoring
Environmental Cleanup: Detection of contaminants, Removal of contaminated materials
Excavation: Grading, Monitoring, Installation of posts, equipment, and materials.
Material Handling: Lifting, Loading, Carrying, and Docking along horizontal or angled surfaces.
Dynamic Sensor Positioning: following a vehicle throughout large areas to maintain local sensing between the suspended platform and vehicle.
Nuclear Waste and Contaminant Management: Inspection of stored waste and contaminated areas, recovery of contaminated material, Field monitoring and inspection.
Toys/Entertainment: Cable-drive, small model or person manipulation and positioning.
Building Surface Access: Panel insertion, Window washing, Material handling along vertical, angled, and/or non-flat surfaces.

VII. Conclusion

LOMAF prototypes provided useful cable tension, work volume and platform control information. The Case 1 configuration has limited work volume since no or limited lateral force is present as the platform approaches a mast or field edge. More efficient work volume, platform leveling and positional control is provided through Case 2 and 3 LOMAF configurations which use downhaul cables to counterbalance the upward tension of the lift cables. These cases improved control substantially because they add lateral platform forces. Cases 2 and 3 increased work volume over Case 1 with four added winches and cables. Case 2 provides an ideal nearly 100% work volume to MSD, but at an even higher cost of added running lines to Case 3.

To achieve similar Case 1 work volume as with Case 2, the MSD must be increased by approximately 30% to 50%. Our preliminary results suggest that the Case 1 LOMAF has utility for applications where a level-controlled crop sensor package can be mounted to the platform, the platform is low in mass, and platform pitch and roll and large MSD versus work volume are permitted. Further cable tension and mast height experiments for all three cases are planned with current and even larger prototypes.

VIII. References

[1] Albus, J. S., Bostelman, R. V., Dagalakis, N. G., "The NIST ROBOCRANE, A Robot Crane", Journal of Robotic Systems, July 1992.
[2] Bostelman, R.V, Albus, J. S., et. al., "Applications of the NIST RoboCrane," Proc. International Symposium on Robotics and Manufacturing, Maui, HA, April 1994.
[3] Dagalakis, N. G., Albus, J.S., Bostelman, R.V., Fiala, J., "Development of the NIST Robot Crane Teleoperation Controller," Proc. Fifth Topical Meeting on Robotics and Remote Handling, Knoxville, TN, April 1993.
[4] Shendure, J., and H. Ji. 2008. "Next-generation DNA sequencing," Nature Biotechnology 26:1135-1145.
[5] Montes, J.M., A.E. Melchinger, and J.C. Reif. 2007. "Novel throughput phenotyping platforms in plant genetic studies," Trends in Plant Science 12:433-436.
[6] Colaizzi, P.D., E.M. Barnes, T.R. Clarke, C.Y. Choi, P.M. Waller, J. Haberland, and M. Kostrzewski. 2003. "Water Stress Detection Under High Frequency Sprinkler Irrigation with Water Deficit Index," Journal of Irrigation and Drainage Engineering 129:36-43.
[7] Wire rope chart - 6x37 & 6x19, IWRC, EIPS rope, http://www.airwinch.com/tools/wireropechart.htm , 2010.

A New Mechanism and Wireless Control System for 7 DOF Camera Robot

Lei Zhang, Shitian Yan, Zhixin Chen and Jia Wang
College of Electronics and Information Engineering
Beijing University of Civil Engineering and Architecture
1st Zhanlanguanlu Road, Xicheng Strict, Beijing, China, 100044
leizhang@bucea.edu.cn, yshitiann@163.com, chenzhixin@bucea.edu.cn, wangjia@bucea.edu.cn

ABSTRACT

High mobility and flexible are important characteristics for camera robots. This paper presents a new robot system 7 DOF that could achieve both of these two abilities. A special mechanism of 7 DOF which allow the robot to work flexibly in variable workspace is designed. The performance of the proposed mechanism is verified reasonable by simulation in matlab. Another special characteristic is the wireless communication module which ensures the robot move without the limits of the wires.

Keyword: Mobility, Flexible, Mechanism, Simulation, Kinematics, Wireless

1. INTRODUCTION

Camera robots are used to carry camera to take videos following a given trajectory. It fits to get some videos with the same mobile viewpoint track according to the director. These videos are used in film post-production synthesis. This work nearly cannot be completed by the cameraman because the human can't guarantee to take photos in the same place with the same velocity many times.

To get videos with high quality, especially for film making, this kind of robot should have characters as follows: 1) Little limited scope of the view track, some film directors may ask aesthetic trajectory for the composition need, they care little about the mechanism practicability; 2) Smoothly moving, motion quality is important in film making, so that the video is stable, and more critical, most professional video camera weight more than 3kg; 3) Little wire constraints, wire of the robot must never display in the camera view field; 4) Indoor and outdoor working place.

In recent years, many successful camera robots have developed. These robots can be divided in two categories depending on mechanism and communication, say a) robots with small size, light weight and low power consumption. b) Robots with excellent mechanism and are suspended and moved by wires. The former category possesses high flexible, but the accuracy is poor. The latter type focused on the structure, but the mobility is limited. Therefore, it is difficult to achieve both flexible and mobility in the traditional way.

In order to ease these problems and optimize the motion track, especially satisfy the velocity and viewpoint requirements, this paper presents a camera robot mechanism. A special 7 DOF including a 3 concentric shaft mechanism was adopted. Flexible adjustable fixed viewpoints could be gotten by the 3 concentric shaft mechanisms. Other series 3 axis mechanisms could take the camera to widen view scope. In additional sliding orbit was discussed in this paper. Sliding orbit with high precision widely expands the video scope. Preferred it is used in a dynamic task and principle sample experiments have proved the effectiveness of such mechanism.

Motion constraint by the wire is a serious problem of this robot. Excessively emphasizing mechanism innovation to solve this problem can only make the robot more complex and difficult to be controlled. A wireless communication pattern is developed to transfer controlling command and joint status between the control center and every distributed motor controller to minimize the motion constraints by communication wire. The communication card has been developed and real time control has been achieved on this communication card. What's more, the online experiments have proved that it could be used in the joint control, even in outdoor environment.

The paper is organized as follows: In section II, detailed introduction in mechanical design is described. Section III mainly discusses the kinematics model of the robot. Section IV depicts the simulation to verify the proposed structure. Section V presents the control system of the robot. Conclusion is made and future directions are pointed out in section VI.

2. 7 DOF MECHANISM

The mechanical design for the 7 DOF camera robot were broken into three main components. The first component is the 3 axis serial mechanism likes the arm of the robot, which can point to every direction by rotating. The second one is the 3 concentric shaft mechanism which plays the role of the robot's hand. The great benefit of this mechanical is the flexible which can achieve. The third component is the sliding mechanism which is the foot of the robot, when the workspace is too big, the foot can help the robot to move instead of standing on the same place. The following sub-sections describe the design concepts of each component.

1) 3 axis serial mechanism

One of the most important tasks for camera robot is that the operators can control the robot to take photos in every direction he needs. This requirement calls for an excellent structure. For that, a series 3 axis mechanism is built

which consist of three serial levers and three degree of freedoms. The first DOF can rotate by y-axis, which can bring the next two levers to move in xoz plane. The axis of L2 and L3 is parallel so that L3 can be up and down when L2 works and L3 can also move by itself. Through the joint motion of L1, L2 and L3, the robot can reach every direction in a wide space. For L3, it is not a straight lever because it can add some other special mechanism which is the 3 concentric shaft mechanism.

2) 3 concentric shaft mechanism

3 concentric shaft mechanisms are well known for its capability of the flexible motion in robot technique. In our system, when the direction has been identified, determining the specific position would be the main problem. In order to ease this problem, the 3 concentric shaft mechanisms become an important design consideration for this camera robot. This 3 concentric shaft structure like the coordinate system, the camera in this system is so flexible that can get to every point in the small space. For that, the operator can easily control the robot to the specific location he wants.

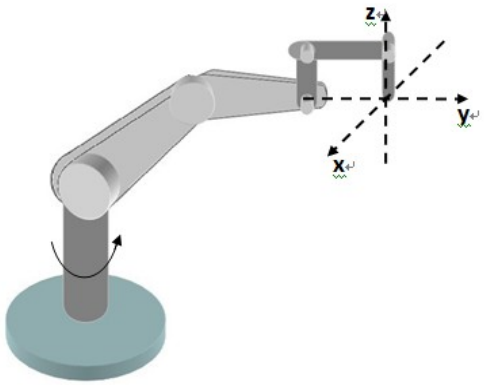


Fig.1 the mechanism of 7 DOF

3) Sliding mechanism

Sometimes the workspace for the camera robot is so big that if you just rely on this six degree of freedoms, you cannot reach the accurate position you expected because the arm of the robot is not long enough. So a sliding module is added for the robot as its foot. When the distance is too far, the robot can first walk there by the sliding module which absolutely widens the robot's workspace.

3. KINEMATICS OF 7 DOF CAMERA ROBOT

7DOF robot is a joint model with six freedom degrees. The joint body is constituted by the rotation of the body structure, big arm, forearm, wrist and other parts. All joints are rotational joints. The origins of the last three axes are intersected at one public point. According to the method of setting up coordinate system, CAC robot's structure and coordinate could be set up as follows:

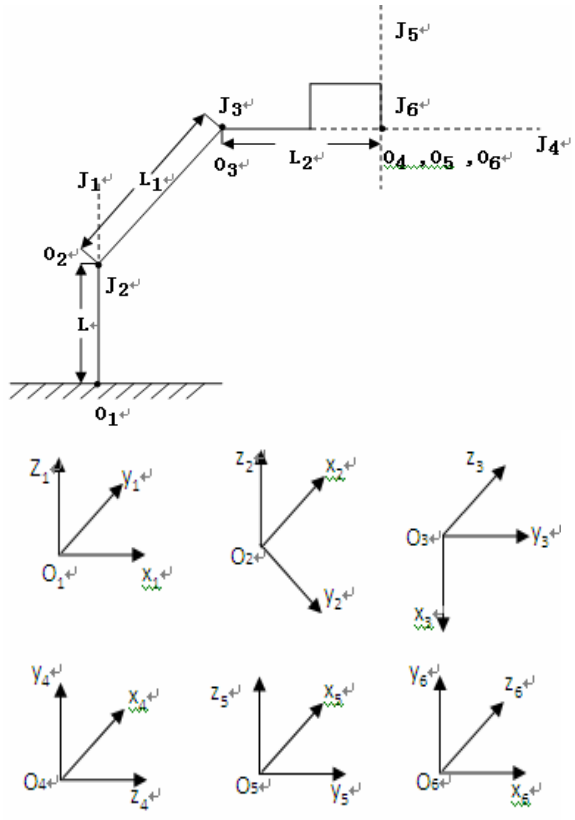


Fig.2 the Coordinate System of 7 DOF

After setting up the coordinate system, the main D-H parameters can be gotten according to the degree of freedom. The parameters are given in table1.

Table 1 the D-H Parameter Table of 7 DOF

Joint rod	Joint angle θ_i	Link offset d_i	Torsion angle α_{i-1}	Length of joint rod a_{i-1}
1	θ1	0	-90°	0
2	θ2	0	0	L2
3	θ3	0	0	L3
4	θ4	0	-90°	0
5	θ5	0	90°	0
6	θ6	0	-90°	0

The kinematics problem for the robot is to give the geometry parameters of levers and the displacement of the joint, finally solve the connecting rod coordinate system.

To solve the kinematics equations, the homogeneous transformation matrix T is often adopted to describe the location and square of the i-frame compared to the frame (i -1).

In the transfer matrix if the link of the D-H matrix were A1, A2… the position and orientation of the end of the robot can be obtained by the following formula:

$$T = A_1 A_2 A_3 A_4 A_5 A_6 \tag{1}$$

To solve the inverse kinematics of the robot, We multiply A_1^{-1} to the left side of the equation (1), and then the other equation can be gotten:

$$A_1^{-1}\,{}^0_6T = A_2 A_3 A_4 A_5 A_6 \tag{2}$$

The left side can be expressed by the following formula:

$$\begin{bmatrix} c\theta_i & s\theta_i & 0 & 0 \\ 0 & 0 & -1 & 0 \\ -s\theta_i & c\theta_i & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & p_x \\ r_{21} & r_{22} & r_{23} & p_y \\ r_{31} & r_{32} & r_{33} & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} = {}^1_6T \tag{3}$$

On the other side, the matrix is:

$${}^1_6T = {}^1_3T\,{}^3_6T = \begin{bmatrix} a_{11} & a_{12} & a_{13} & p_x \\ a_{21} & a_{22} & a_{23} & p_y \\ a_{31} & a_{32} & a_{33} & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}$$

$$a_{12} = c_{23}c_4s_5 + s_{23}s_4s_5$$

$$a_{13} = -c_{23}c\theta_4c\theta_5s\theta_6 - c_{23}s\theta_4c\theta_6 + s_{23}s\theta_4c\theta_5s\theta_6 - s_{23}c\theta_4c\theta_6$$

$$a_{21} = s_{23}c\theta_4c\theta_5c\theta_6 - s_{23}s\theta_4s\theta_6 + c_{23}s\theta_4c\theta_5c\theta_6 + c_{23}c\theta_4s\theta_6$$

$$a_{31} = -s_5c_6$$

$$a_{32} = -c_5$$

$$a_{33} = s_5s_6$$

$$p_x = c_{23}L_2 + c_2L_1$$

$$p_y = s_{23}L_2 + s_2L_1$$

$$p_z = 0$$

In these two matrices, eq.3 and eq.4, the corresponding items are equal. Using these equations, all the joint angles can be calculated.

4. SIMULATION IN MATLAB

In order to prove the validity of the proposed robot's mechanism, simulation is an important method which can be considered. The robot toolbox in matlab provides an excellent environment to simulate many situations. The kinematics simulation of robot is divided into three stages: The establishment of the robot's leverages; kinematics simulation and inverse kinematics simulation. By analysis the simulation, the mechanism is proved reasonable.

1) The establishment of the robot's leverages

To establish the leverages of the robot, the D-H parameters are necessary which have been analyzed in the table before. Using these parameters, the Three-dimension leverages' picture can be displayed in the space coordinate system by programs. The leverages of the robot are shown in Figure. In slide mode, you can manually adjust each joint, and then observe the trajectory space of robot.

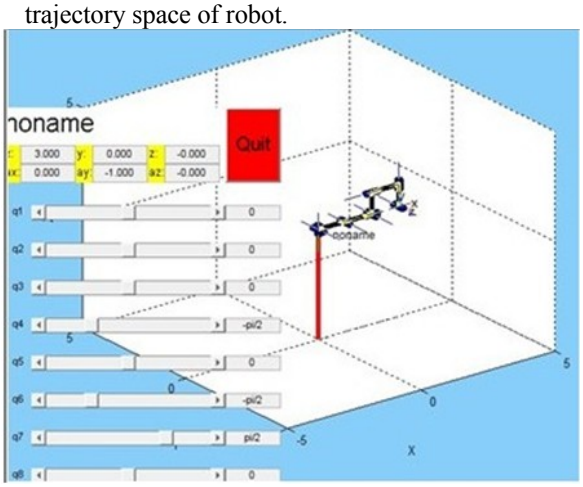


Fig.3 the Three-dimension Image and Slider-controlling Picture

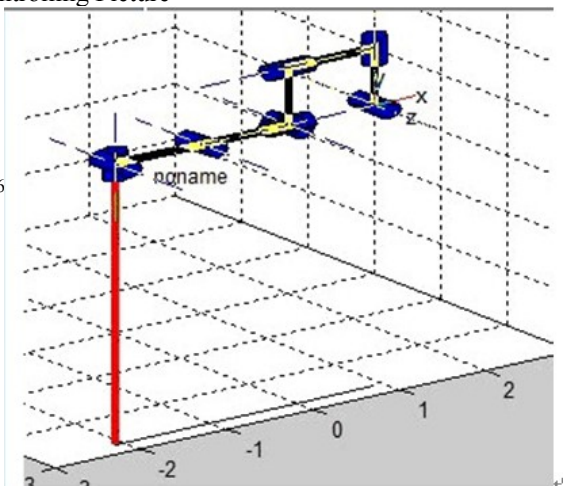


Fig.4 the Three-dimension Image

2) The Kinematic Simulation of Robot.

If the first joint angle and the last joint angle of the lever are given, the transfer matrix of every degree of freedom can be easily calculated in matlab. Just by programs, the whole process of kinematics can be seen. In Robotics Toolbox, fkine ( ) and jtraj ( ) could be used to calculate Joint coordinate trajectory and transformation matrix of robot's movement. For example, if the first joint angles are qz = [0 0 0 0 0 0 0 0] and the last joint angles are qzn=[0.5 1 2 1.2 3 1 2 2.2], after 14 steps, the robot can complete the motion and the transfer matrixes can be generated in matlab. The following is the calculation. Figure5 shows the angles.

The first transfer matrix is:

$$\mathrm{T}(:,:,1) = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The last transfer matrix is

$$T(:,:,14)=\begin{bmatrix} -0.3769 & 0.4139 & -0.8286 & -0.0437 \\ -0.2928 & -0.9020 & -0.3176 & 1.1042 \\ -0.8787 & 0.1230 & 0.4612 & -0.6153 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

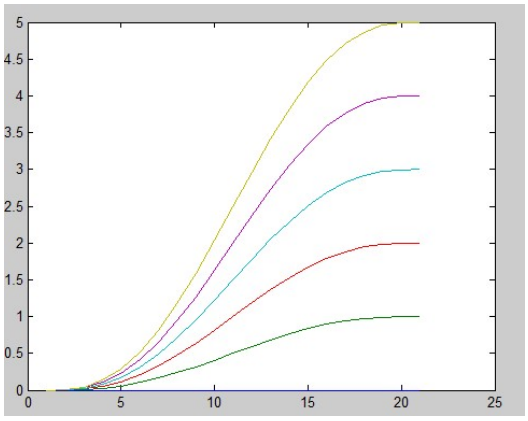


Fig.5 The trajectory of angle θ_1~θ_6

In the robot simulation, paths or trajectories frequently need to be deal with. So we can generate the picture of the joint angles using the robotic toolbox and then get the joint speed and joint acceleration. The following pictures show the parameters of joint 2.

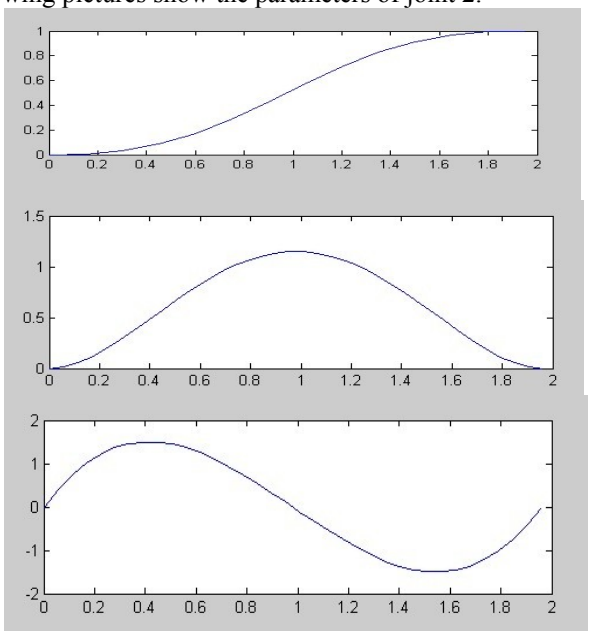


Fig.6 The second joint angles, speed and acceleration in space trajectory

3) The Inverse Kinematic Simulation of Robot.

The same as the kinematic simulation, ikine ( ) and ctraj ( ) command in Robotics Toolbox could be used to calculate the trajectory. The trajectory planning of robot mainly is divided into two types; in this paper the trajectory planning of point to point (PTP) is adopted to explain this theory. First define the Initial point of the T matrix T1 = [1.000 0.000 0.000 3.000; 0.000 -1.000 0.000 1.000; 0.000 0.000 -1.000 -1.000; 0.000 0.000 0.000 1.000], then define the nth point of the T matrix T2 = [-0.9459 -0.2137 -0.2441 1.5028; 0.2966 -0.2649 -0.9175 -0.0846; 0.1314 -0.9403 0.3140 -2.0148; 0.000 0.000 0.000 1.000]. So the trajectory can be calculated. Because there are many paths from one point to another, the calculation may be not only one. Table2 are shown the calculation.

Table 2 Angle changes of 7 DOF

θ / DOF	1	2	3	4
1	-0.0000	-0.0161	-0.0227	-0.0273
2	0.0000	-0.1106	-0.1590	-0.2003
3	0.0000	0.0767	0.1286	0.1807
4	-0.0000	0.2555	0.3817	0.4891
5	-0.0000	-0.4183	-0.7406	-1.0433
6	-0.0000	0.0104	0.1098	0.2273

5. WIRLESS ROBOT CONTROL SYSTEM

In traditional robot control system, serial communication module, professional bus communications, or industrial Ethernet communications are always used to transport dates. Although these communication modes are stability, the expansion is too weak and the mobility is poor. You at least need to consider how to connect these lines to the mechanism, which is a complex work you cannot avoid. When the system is in trouble, it is difficult to find out the fault point. To solve these problems, a wireless communication module is adopted to establish a short-range wireless communication network. The advantage of this system is that you need not to consider the lines at all.

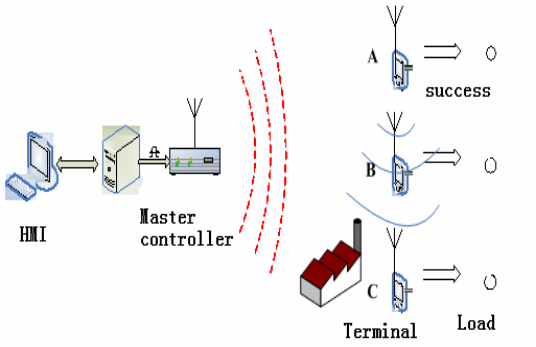


Fig.7 the communication

In this robot system, the wireless serial communication card is used to send or receive the information between the robot and the control center. The multiple control information is transferred by the special network. But finally they must be used in the robot control drivers. In this system, the wireless transport distance is covered about 50m. The schematic is given as Fig.7.

In the real control work, the desirable direction and angel value are easy to input to the system by a joystick.

The angle data are detected by the sensors equipped on the joystick, and communicated to the teleportation system by USB interface.

6. CONCLUSION

This paper mainly describes the design of 7 DOF camera robot that overcomes the limitations of flexible and mobility. A special mechanism is adopted in this system and by analyzing the kinematics of the robot the structure is verified to be reasonable. What's more, the wireless serial communication card is exploited in the robot control system, which can allow the robot to work in variable environments without the wires limitation. There are two directions for the next step of the research:

(1) Make some experiments to prove the accuracy of the wireless serial communication card and put it into practice.

(2) Optimize the trajectory and velocity which allow the robot to adapt more different missions.

7. ACKNOWLEDGEMENTS

I would like to express my gratitude to all those who helped me during the writing of this thesis. My deepest gratitude goes first and foremost to general project of Beijing Municipal Commission of Education (KM201010016006). Without its support, this thesis could not have reached its present form.

8. REFERENCE

[1] GAO Dao-xiang, and XUE Ding-yü, Simulation Research of Robust Adaptive Control System for Robotic Manipulators Based on MATLAB/Simulink[J] Journal of System Simulation. 18 (7), 2006.

[2] Curtis W. Nielsen, Michael A. Goodrich, and Robert W. Ricks, Ecological Interfaces for Improving Mobile Robot Teleportation[J], IEEE Trans. On Robotics, 23(5):927-041, 2007.

[3] Yin ji-ying, He li-ting, Analyzing and resolving of Motoman — HP6 robot kinematics and dynamic [J] Machinery Design & Manufacture. NO.3, March, 2009.

[4] Xin ling, Gu hong-bing, Wu dong-su, MATLAB simulation of an articulated two-link manipulator[J] Instrument User. 16 (2) 2009.

[5] Cheng yong-lun, Zhu shi-qiang, Luo Li-jia, Liu song-guo, Kinematics analysis and simulation of QJR welding robot based on Matlab, mechanical & electrical engineer magazine，24 (11), 2007.

[6] Schilling K, Jungius C. Mobile robots for planetary exploration [A]. 2nd IFAC Conference on Intelligent Autonomous Vehicles [C]. Espoo, Finland: Pergamon Press Ltd, 1995. 109 - 119.on Intelligent Robots and Systems, Oct 13-Oct 17, 2006, 1190-1195.

[7] Eddings Richard, Horne Tim, Ledbetter James, Pustka Brian. Rover subsystem TSGC lunar rover design project [R]. 1996.

[8] Cox I J. Blanche an experiment in guidance and navi gation of an autonomous robot vehicle [J] .IEEC transaction on robotics and Automation. 1991, 7 (2):193 - 203.

[9] Dollarhide Robert L, Agah Arvin. Simulation and control of distributed robot search teams [J]. Computers and electrictity Engineering, 2003, 29 (5):625 - 642.

[10] Zhao Qingjie, Sun Zengqi. Image-based robot motion simulation [J]. Optics Communications, 2002, 205 (4 - 6) :257-263

[11] Peter I Corke. Manual of Robotics TOOLBOX for MATLAB (Release 6) [EB]. Http: / /www. Cat. Csiro. Au /cmst/ staff /pic /robot. Html, 2001 - 02 - 11.

The Case for Millimeter Wave Power Beaming

Narayanan Komerath
School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
Padma Komerath
SCV Inc, Johns Creek, GA 30022, USA

Abstract

Millimeter waves are used primarily for imaging and short-range signal transmission. One cited advantage is that many frequencies in this range get attenuated within a short range, minimizing interference between devices. In this paper arguments are laid out why power beaming through the atmosphere using selected frequencies in the millimeter wave regime will enable breakthroughs in power delivery to and from off-grid areas, synergy between terrestrial and space-based power generation, and an eventual progression to space solar power. The arguments against using millimeter waves are reviewed and compared against the implications of recent developments. The paper assesses the impact of millimeter wave beaming on retail power beaming as well as on the architectures for large-scale intra-national and intercontinental power exchange and power delivery from space. Wavelengths near 100 GHz are suitable for terrestrial short-range beaming, while those near 220GHz are suitable for transmission via stratospheric platforms and space. Advances in optical heterodyning, evaporation ducts, absorption by atmospheric water, oxygen and nitrogen, and synergy with photovoltaic and mobile telecommunications infrastructure are relevant to bringing about millimeter wave power beaming solutions.

Introduction

Since the days of Tesla [1], the idea of wireless power transfer has been considered for many applications. Present interest comes from 3 reasons:

1. Power delivery to and in off-grid communities.
2. Power transfer to and from islands.
3. Power transfer through and from Space.

This paper begins by reviewing the tradeoff between frequency, infrastructure size and efficiency, and why it is essential to consider the millimeter wave regime at 220 GHz and 100 GHz. The rest of the paper reviews work done on the generation, reception, transmission and attenuation of electric power in the millimeter wave regime.

The dream of Space Solar Power

While proponents of terrestrial nuclear fusion power plants argue their dream of unlimited inexpensive and clean power, it is useful to point out that we already get nuclear fusion power in plenty from the Sun, and without the liabilities of terrestrial nuclear plants. Converting sunlight to electric power on earth is immediately practical. However, if the conversion is done in space, the residual heat from the conversion process (60 to 80% of the incident power), stays in Space rather than heating up the atmosphere. Nuclear plants achieve the highest "capacity factor" of all terrestrial power plants, being available to operate at full rated power 85 to 90% of the time, year-round. On the other hand, a space solar power plant will operate essentially 100% of the time at a steady power level, and can deliver power anywhere on Earth. This is the dream of Space Solar Power (SSP) Reasons why it has remained a dream are seen below.

In the early 1960s, the first satellites had just been launched, and the space program was racing forward. Arthur C. Clarke [2] had popularized the unique properties of the Geostationary Earth Orbit (GEO) at 36,000 kilometers above the Equator, where a satellite would appear to be stationary with respect to the spinning Earth. Several exciting applications were proposed with spacecraft placed in GEO, including the communication satellites that have since become the largest manifestation of Space based commercial activity. Peter Glaser [3], then a Vice President of the Arthur D. Little Company, presented a concept for large platforms in GEO where solar power would be collected, converted to electric power, and transmitted to receivers on Earth as microwave beams. Note that the laser[4] had only been invented in 1959, while microwave beaming had been in use since World War II. The US National Aeronautics and Space Administration (NASA) studied the concept [5] in consultation with the US Department of Energy, (DOE) which ultimately was given responsibility for developing SSP. By 1978, a joint DOE/NASA Reference SSP System definition included environmental and societal impact reports.

Several design choices of that era still remain, and must be reviewed. GEO is high enough (orbit radius >> Earth radius) that a few satellites there would ensure continuous visibility of the Sun. The geostationary aspect was a unique advantage [3]. The distance from Earth to GEO is however large. This can be seen from Equation 1, relating the antenna sizes for full beam capture to the wavelength and the distance. D_t, D_r are the diameters of the transmitter and receiver, while λ and S are the wavelength and beaming distance respectively.

$$\frac{D_r D_t}{\lambda S} = 2.44 \qquad \text{.................................(1)}$$

At microwave frequencies (e.g. 2.45GHz), even the primary core of a Gaussian beam coming from GEO, containing some 84% of the beam energy, would be on the order of hundreds of kilometers across, even with kilometer-sized transmitters in GEO. The huge power level required to justify such a receiver would require a high-capacity power grid connected to it. Once built, such large stations must operate without much interruption to be viable. This drove the choice of a frequency well below the 10GHz frequency that excites water molecules, to keep it as immune as possible to the weather. The receiver size is still so large that designers limited themselves to capturing only 84% of the beam, allowing the rest to spill over. Thus the transmission efficiency is effectively less than 80%. The receiver area for over 98% beam capture is just over 1.5 times that for 84% capture, increasing the diameter 1.23 times.

Terrestrial power generation is over 4,000 GigaWatts electric (GWe). A simple calculation shows the size of the satellite infrastructure required to deliver 1GWe to Earth. The AM0 (air mass zero) solar spectrum at Earth's orbit around the Sun delivers roughly 1.38 kilowatts (KW) of solar power per square meter. With a conversion efficiency of sunlight to DC using photovoltaic (PV) solar cells of 25%, DC to beamed microwave conversion of 80%, 95% atmospheric transmission, 80% beam capture at the surface and microwave to DC or AC conversion of 80%. Thus what is delivered to the terrestrial grid is only 12.6% of the solar power captured in orbit. The capture area needed to deliver 1 GWe at Earth is thus nearly 6 million square meters, or a square 2.44 kilometers on each side. The transmitting antenna also has to be very large, per Equation (1). Clearly this poses a huge problem of delivering materials to, and constructing this large station in GEO.

Numerous studies have been conducted on SSP in the decades since the 1960s [6-12]. They have all been focused on a GEO-based architecture. Japanese researchers [12-16] proposed a demonstration SSP craft in Low Earth Orbit (LEO, a few hundred kilometers above Earth), beaming at 2.45GHz over a broad footprint. They argued that even the very low intensity captured using portable antennae would help explorers and residents in remote areas charge their communication devices and obtain emergency power. This is an important point: when one has no power grid, the cost per unit power or energy is extremely high, and the first few watt-hours of power are precious for communications, navigation and life-saving applications. Bekey and Boudreault [17] proposed a route to quick revenue generation by beaming power from Canadian power plants to an antenna in a mountain lake in Japan via a GEO relay satellite. The water would serve to store spillover energy as heat. European interest, driven by greater sense of urgency, has shifted to large terrestrial DC grids connecting solar and wind power plants [18-20]. The point above is that the choice of GEO leads to prohibitive size and cost of both the Space portion and the Earth portion of SSP. Thus we conclude that beaming using low frequencies from GEO will not work for large-scale implementation of Space Solar power.

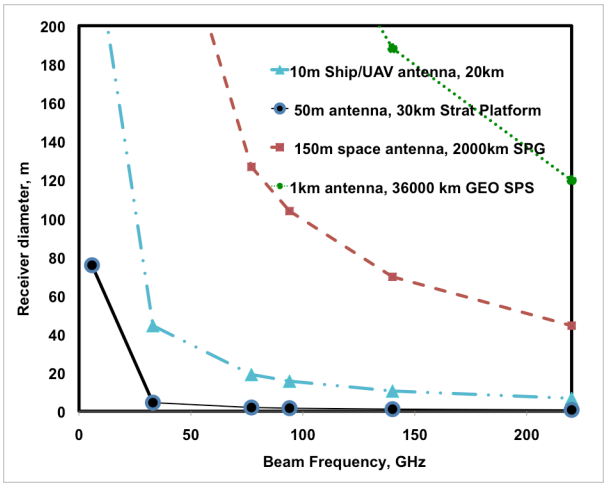


Figure 1: Receiver size on earth versus beam frequency and distance

Millimeter Wave Beaming

The simple answer to the above problem, from Equation (1) is to shift to a much higher frequency. Lasers would be excellent if possible to generate with good efficiency and low mass per unit power, and are being investigated by some researchers. Currently they pose severe terrestrial restrictions due to radiation safety concerns, and Space restrictions due to the prohibition on weapons in Space. In the millimeter wave regime, there are good atmospheric propagation windows near 100 GHz, 140 GHz, and especially at 220GHz. Secondly, the GEO choice can now be abandoned, and beaming can be done from much lower orbits, thereby decreasing the antenna sizes by another factor of up to 20. Figure 1 compares receiver antenna sizes for several architectures. The GEO-based architecture is far off this chart. The reasons in favor of considering millimeter wave beaming from lower orbits are summarized below:

1. Reduction in antenna area by 2 orders of magnitude compared to the 2.45-5GHz regime.
2. Unlike the 1960s, when the antenna had to be aimed mechanically, unlike the Phased Array beam steering technology of today.
3. The substantial problems of accurate pointing over large distances have been solved in the beam weapons community and the space program. In addition, adaptive control can iteratively refine the beaming accuracy before full-power operation. Today "smart antennae" use Digital Signal Processing (DSP) to identify moving cellphone locations in real time to minimize cross-talk [21].
4. The fears about misdirected death rays are misplaced. The full beam operates only when a trigger link is operating.
5. The above considerations make it possible to consider

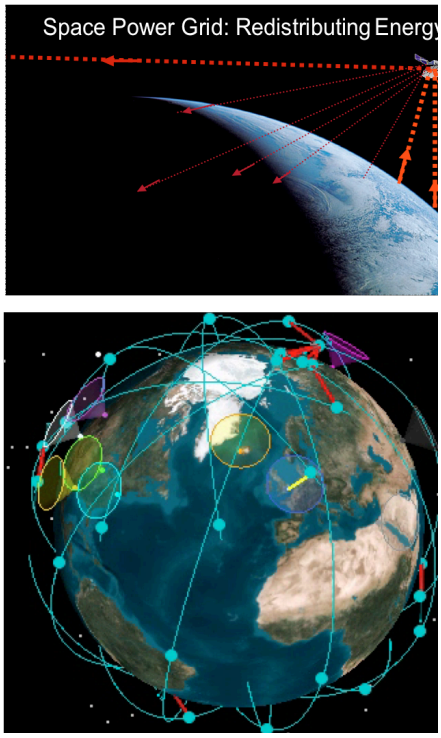


Figure 2. A LEO constellation can serve to exchange power between solar plants on Earth

beaming from Low or Mid Earth Orbit (L/MEO), about 2000 kilometers high, rather than the 36000 kilometers from GEO. This further reduces antenna size.

6. With L/MEO satellites, transient passage over any ground station, and several space-to-space passes to reach different places on Earth are issues. Space-space power exchange is also needed for any scaled-up SSP system.
7. In the 1960s, millimeter wave power was generated using vacuum tubes and gyratrons, at a very high cost in mass per unit power, and severe maintenance demands. Today this situation has improved dramatically, with very large arrays of mass-produced solid-state devices [22]. These devices also achieve higher conversion efficiencies from DC or AC to millimeter waves, so that the heat dissipation problem is reduced substantially.
8. Frequency selection was a difficult issue with vacuum tubes. Today, with solid state devices, digital signal processing (DSP) and phase-locked loop (PLL) technology enable quick changes and accurate choices to narrow frequency bands. Conceivably, the same array could operate at more than one frequency, or at multiple frequencies intermeshed. This has many implications. There are large issues, for instance whether this works when the purpose is power conversion rather than generating a small-amplitude signal.
9. One of the most interesting features of beaming from a low orbit constellation is that power can reach every point on Earth's surface, unlike GEO-based craft which would be too low on the horizon at high latitudes. This can be done without massive terrestrial wired grids. Hence it makes sense to use frequencies that enable the location of small receivers in remote places and on islands that are inaccessible to the wired grids.
10. With a wireless power infrastructure, one can consider integrating a huge number of standalone power generating devices into a wireless grid. This opens up the possibilities for micro renewable power generators worldwide.
11. Automotive radar at 77 and 79 GHz, and some quasi-weapons in the 80-94GHz range are examples, the latter of higher-power applications.

TABLE 1: 1-way atmospheric Attenuation of Micro and millimeter waves at 45 degrees azimuth [23].

Frequency	220GHz	140GHz
Dry atmosphere	0.8dB (17%)	0.8dB (17%)
Medium humidity	3dB (50%)	1.0dB (20%)
High humidity	4.6dB (65%)	1.4dB (28%)

Present Scope and Objectives

The above logic drives this paper. Past research by the authors and co-workers [24] has established the Space portions of a solution to the Space Solar Power (SSP) problem, and the opportunity for synergistic, evolutionary development with terrestrial renewable power generation. Figure 2 shows the Space Power Grid concept. At the other end of the scale, the possibility of synergy with micro renewable power generation and distribution, and with the mobile telephone and railway communication networks, has been explored and some general guidance established [25]. In this paper, we are continuing to refine the parameters for millimeter wave beaming, with respect to atmospheric beaming and waveguides. There are daunting obstacles to implementing a millimeter wave power grid, integrating terrestrial and space-based sources and customers. Below we list some of these obstacles, along with potential solution approaches.

Survey Of Issues And Solutions

1. Expense of solid state arrays vs. PV arrays

Millimeter wave power generation using solid state devices is still quite expensive. Transmitters and receivers may need thousands or millions of individual solid state devices, making their costs comparable to those of solar photovoltaic arrays. One might well ask why one should go to the complexity of beaming, rather than simply buy photovoltaic arrays.

One answer is that such receiver arrays may well be used for multiple purposes, receiving highly intense electric power when it is beamed, but also converting solar power directly when it is available [26]. This would improve the value of the infrastructure used to build these devices. There is increasing interest in co-locating receivers with solar photovoltaic (PV) arrays [27-31].

2. Atmospheric Propagation Losses

Papers in the 1980s considered 220GHz beaming to military satellites, as well as beaming through the mushroom clouds of nuclear explosions for damage assessment. The atmospheric propagation of millimeter waves has been studied by the military [31] and by astronomers for imaging purposes. Such spectra are generally published for the dry atmosphere above high-altitude locations of specific observatories. As Figure 3 shows, the "window" below 10 GHz is essentially transparent. Other windows occur near 100 GHz, 140GHz and 220 GHz. The ones near 140 and 220 GHz are similar in that roughly 10% loss is encountered in dry atmospheres above 2000 meters, and that in even moderate rain, both suffer high levels of opacity. Table 1 from Ref. [24] captures the effects of precipitable moisture for oblique propagation through the atmosphere, and suggests that 220GHz is essentially as good as 140GHz for transmission, and equally bad when there is rain or snow. However, the same report also shows that for wide swathes of the United States, substantial precipitation is typically less than ten hours per year. Similar studies in other nations will identify good locations for ground receivers connected to the grid, that can be used when other places are under bad weather. The

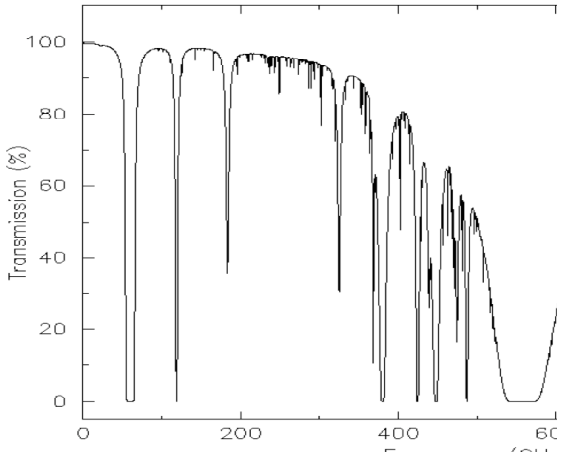


Figure 3: Transmission through a dry atmosphere (less than 50mm precipitable moisture). Astronomical observatory data, Mauna Kea, Hawaii.

implications of sending very high power beams through moist atmospheres remain to be understood. Past work has investigated "evaporation tunnels" at low altitudes which are relevant to this issue.

Horizontal propagation at low altitudes is a much worse problem. Figure 4 from Reference [32] shows that attenuation decreases rapidly as altitude increases, so that propagation should occur as much as possible at higher altitudes. Line 'A' shows average absorption at sea level, 293K, with 7.5grams per cubic meter of water vapor. Line B shows attenuation in dB per kilometer at 4000 meters pressure altitude, at 273K, with 1 gram per cubic meter of water vapor. It is seen that there is a substantial advantage to staying near 100 GHz rather than 220GHz for this application. Figure 5 shows the effects of atmospheric moisture on attenuation. Here the basic message is that the attenuation is unacceptably high, except in the very low frequency end of the spectrum. However, the experience of satellite television dish owners suggests that even in the regime below 10GHz, performance during rainstorms is often poor. Hence, going around, or burning through heavy rain are the best alternatives.

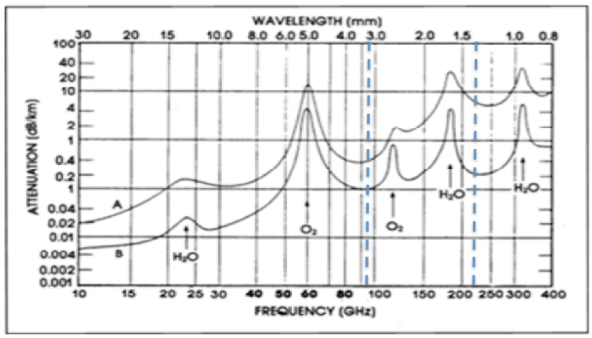


Figure 4: Attenuation of horizontal millimeter waves at two altitudes. From Ref. [33]

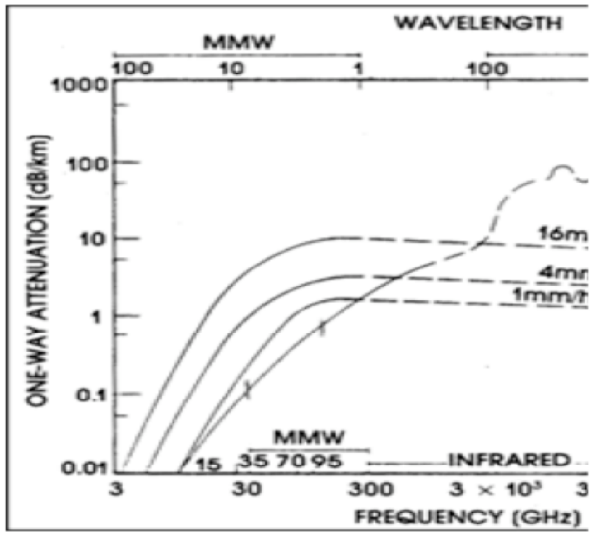


Figure 5: Effects of atmospheric moisture on horizontal attenuation.

3. High Beam Intensity

Small antenna size amplifies another problem: Delivering substantial power requires very high intensity. Atmospheric transmission at such power levels will require well-considered changes in safety regulations, and these will in turn require technical innovations and multiple layers of protection. The intensity is likely to be very high only for the ground to space transmission stage from a power station, and for the space-to-Earth stage, where the receiver is best located on a high-altitude platform for further retail distribution to receivers on the ground. At these locations, the high intensity can be handled by appropriate thermal management systems, or simply by using large collectors.

4. Health Effects of Millimeter Waves

Some scattering is inevitable near receivers and transmitters, and the health implications of these frequencies must be studied. Research in this area is being given a boost since airport security systems started using millimeter wave scanning of passengers, and frequent flyers such as airline crews started expressing concerns about radiation exposure.

4. Generation of Millimeter Waves

The biggest obstacles to the use of millimeter waves in power transmission are the cost and low efficiency of methods to generate them. Only optical / photonic heterodyne methods of generation are expected to yield high efficiency. Conversion efficiencies as high as 50% are expected when beat frequencies between laser beams are used to generate millimeter waves. Phase noise is reduced using an optical phase-locked loop [33-35]. While this is still quite inefficient, it opens the way for future schemes where broadband sunlight is used to generate millimeter waves directly. This is the ultimate promise of millimeter waves as transmission mode for electric power converted from sunlight.

Other issues to be considered include the technology of phase array antennae, tuning and synthesizing to narrow wave bands using phase-locked loops, waveguide technology and small scale testing. These are used to project technology and arrive at reasonable parameters to use in simulating the performance of systems optimized for various portions of the wireless beaming infrastructure. These parameters are essential to refine the projection of end-to-end efficiency.

Results from the SPG Concept Analysis

A concept analysis program has been developed, using simple FORTRAN90 software. This has been used [36] to explore parameter regimes and combinations that will satisfy various constraints. The criterion for viability is the Net Present Value climbing back to zero within a desired number of years at a reasonable rate of return on investment. Each of these parameters depends on the particular circumstances of the system being considered. This process has to-date shown that a Space Power Grid approach where power is exchanged through Space in a first phase, would become self-sustaining and break even at a 6% Return on Investment within the 17-year lifetime of a typical Low Earth Orbit (2000km) constellation that reaches 96 satellites.

We are yet to perform quantitative calculations of the retail end of the millimeter wave beaming architecture. However, some conclusions can be derived logically. The major intercontinental exchange between the daylit and night sides of the globe should

occur between power plants, which have the protected land area to easily place large transmitter/receiver arrays on their property. However, the establishment of such a beamed power infrastructure also opens opportunities for extensive retail beaming. As discussed in [37] the logical optimum architecture appears to be one where major power plants perform exchanges through Space, and also feed the wired power grid. Smaller local plants beam power to stratospheric platforms (large lighter-than air airships), which beam across to other platforms and down to retail customers in remote areas. There may also be a role, as a reviewer pointed out, for regional or local DC grids to exchange power between renewable sources and their customers, similar to the model being discussed by the Europeans. Satellites also beam down to the stratospheric platforms. With the large antennae that can be mounted on these platforms, receivers on the ground can be small enough to be mounted on single-family homes or can be portable.

Conclusions To Date

In this paper, the reasons for millimeter wave power beaming are considered, along with the associated problems, and avenues for their possible solution.

1. Millimeter wave beaming is essential to success in bringing Space Solar Power to earth, because it brings antenna and transmitter sizes down to sizes practical for retail installation and perhaps even portable emergency installations.
2. It enables retail distribution of power to off-grid locations and islands.
3. It enables synergy with distributed micro renewable power architectures.
4. Technical options for building millimeter wave generators, transmitters and receivers are improving rapidly towards mass production
5. Synergy with photovoltaic arrays is feasible and promises substantial improvements in the cost-effectiveness of micro power architectures.
6. Concerns about atmospheric absorption and scattering remain to be resolved, but there are interesting options to pursue.
7. Optimal choices for propagation through moist atmospheres and for long-distance dry air and vacuum beaming differ.
8. Solid-state arrays promise the option of using several beam frequencies from the same arrays in order to serve multiple purposes.
9. The possibility of evaporation ducts for atmospheric propagation should be explored.
10. The optimal architecture appears to consist of a 2000km constellation of 96 power grid satellites performing intercontinental power exchanges, and a network of stratospheric platforms for local and regional exchanges and retail distribution to customers.

Acknowledgements

This first author is supported by the NASA EXTROVERT cross disciplinary learning project. Mr. Tony Springer is the Technical Monitor.

References

[1] Tesla, N., “Transmission of Power”: Polyphase System; Tesla Patents”. Westinghouse Electric and Manufacturing Company, Pittsburgh, PA, 1893, 246p.

[2] A. C. Clarke, "Extra-Terrestrial Relays," Wireless World, Vol. 51, No. 10, pp. 305-308, 1945

[3] Glaser, P., "Power from the Sun: It's Future", Science, Vol 162, pp. 856-861, 1968.

[4] Gould, R. Gordon (1959). "The LASER, Light Amplification by Stimulated Emission of Radiation". In Franken, P.A. and Sands, R.H. (Eds.). The Ann Arbor Conference on Optical Pumping, the University of Michigan, 15 June through 18 June 1959. pp. 128.

[5] Anon, “Satellite Power Systems (SPS) Concept Development and Evaluation Program Preliminary Assessment” DOE/ER 0041, US Department of Energy, Satellite Power System Project Office, NASA TM81142, 1979.

[6] NASA, Sep. 19-20 1995, " Space Solar Power, An Advanced Concepts Study Project", NASA HQ, NASA LeRC SSP Technical Interchange Meeting,", Washington, D.C. 1995.

[7] Potter, S., Henley, M., Davis, D., Born, A., Bayer, M., Howell, J., Mankins, J., “Wireless Power Transmission Optionsfor Space Solar Power”. Proceedings, State of Space Solar Power Technology Workshop, Lake Buena Vista, Florida, 2-3 October 2008.

[8] Anon., “The Final Proceedings of the Solar Power Satellite.” Program Review, NASA-TM-84183.

[9] Stancati, M.L. et al. Space Solar Power, a Fresh Look Feasibility Study, Phase 1, SAIC-96/1038, NASA Contract NAS3-26565.

[10] Henley, M., Potter, S., Howell, J., Mankins, J., “Wireless power transmission Options for Space Solar power” ”, International Astronautical Congress, 2002

[11] Anon., Laying the Foundation for Space Solar Power” (2001) – National Research Council.
http://www.nap.edu/execsumm/0309075971.html.

[12] Saiki, T., Uchida, S., Motokoshi, S., Imasaki, K., Nakatsuka, M., Nagayama, H., Saito, Y., Niino, M., Mori, M., “Development of Solar-Pumped Lasers For Space Solar Power Station”. International Astronautical Congress, Fukuoka, Japan, Oct. 2005.

[13] Sasaki, S., Tanaka, K., Higuchi, K., Okuizumi, N., Kawasaki,S., Shinohara, M., Senda, K., Ishimura, K., “Feasibility Study of Tethered Solar Power Satellite”, International Astronautical Congress, Fukuoka, Japan, Oct. 2005

[14] Kuroda, Y.,Nagatomo, M., Collins, P., "Japanese Perspective on Power from Space for Earth", SPSRio 92 - Space Power Systems and Environment in the 21st Century Symposium, SEE, Paris, 1992. http://www.spacefuture.com/archive/japanese perspective_on_power_from_space_for_earth.shtml

[15] Nagatomo, M., “An Approach to Develop Space Solar Power as a New Energy System for Develop Countries”. Solar Energy, Vol. 56, No 1, pp 111-118.

[16] Nagatomo, M., Itoh, K., “An evolutionary satellite power system for International demonstration in developing nations”. Proc. SPS91, 1991. pp 356-363. Also at http://www.spacefuture.com/archive/an_evolutionary_satellite_power_system_for_international_demonstration_in_developing_nations.shtml Viewed Sep. 12, 2006.

[17] Bekey, I, Boudreault, R., “An Economically Viable Space Power Relay System”, Acta Astronautica, Vol.44, Nos. 7-12, pp.385-390, 1999

[18] Geuder, N., Quaschning, V., Viebahn, P., Steinsiek, F., Spies, J., Hendriks, C., “Comparison of Solar Terrestrial and Space Power Generation for Europe”. 4th International Conference on Solar Power from Space – SPS’04, Granada, June 2004.

[19] Trieb, F. et al, “Trans-Mediterranean Interconnection for Concentrating Solar Power. TRANS-CP. Final Report, German Aerospace Center (DLR), June 2006. http://www.dlr.de/tt/trans-csp

[20] Summerer, L., Ongaro, F., "Solar Power From Space: Validation of Options for Europe". Paper ACT-RPR-NRG-2004-ESA, European Space Agency, 2004.

[21] Cooper, M., "Antennas Get Smart". Scientific American, July 2003, p. 49-55

[22] Pham, A.V.H.; Laskar, J.; Krishnamurthy, V.B.; Cole, H.S.; Sitnik-Nieters, T., "Ultra low loss millimeter wave multichip module interconnects". IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part B: Advanced Packaging, Vol. 21, Issue 3, Aug. 1998, p. 302-308

[23] Anon, "Millimeter Wave Propagation: Spectrum Management Implications," Office of Engineering and Technology, Federal Communication Bulletin No. 70, Washington DC, July, (1997).

[24] Komerath, N.M. and Komerath, P.P., "Implications of Inter-Satellite Power Beaming Using a Space Power Grid". Proceedings of the IEEE/AIAA Aerospace Conference, Big Sky, Montana, March 2011.

[25] Komerath, N., Chowdhary, G., "Retail Beamed Power for a Micro Renewable Energy Architecture: Survey". Proceedings of the ISED/SEEDS Conference, Bhubhaneshwar, India, December 2010.

[26] Er, S., Yegin, K., "2.4GHz Antenna Integrated Solar PV Cell". Engineering Project Report, Yeditepe University, Faculty of Engineering and Architecture, Department of Electrical and Electronics Engineering, Istanbul, 2009 http://www.emo.org.tr/ekler/3cd3d45725f43cb_ek.pdf

[27] Drapalik, M., Schmid, J., Kancsar, E., Schlosser, V., Klinger, G., "A Study of the Antenna Effect of Photovoltaic Modules". Paper 333, International Conference on Renewable Energies and Power Quality, (ICREPQ'10), Granada (Spain), 23rd to 25th March, 2010

[28] Costia, D.I., Popescu, <.O., Popescu, C.L., Craciunescu, A., "Photovoltaic Solar Cell Like Receiver For Electromagnetic Waves in VHF-UHF Bands". Paper 303, International Conference on Renewable Energies and Power Quality (ICREPQ'10), Granada (Spain), 23rd to 25th March, 2010

[29] Henze, N.; Bendel, C.; Kirchhof, J., "Photovoltaic Power Supply and Antennas in one Device for Wireless Telecommunication Equipment". Telecommunications Conference, 2005. INTELEC '05. Twenty-Seventh International, Berlin, Sep. 2005, p. 71-76

[30] Henze, N.; Giere, A.; Fruchting, H.; Hofmann, P.; "GPS patch antenna with photovoltaic solar cells for vehicular applications", Vehicular Technology Conf., 2003. VTC 2003 vol 1, pp:50 – 54.

[31] Liebe, H.J., and Hufford, G.A., "Modeling Millimeter Wave Propagation Effects in the Atmosphere". AGARD Report 454, North Atlantic Treaty Organization, 1989.

[32] US Department of Transportation.
http://ops.fhwa.dot.gov/publications/viirpt/sec5.htm

[33] Gliese, U, Nielsen, T.N., Nørskov, S., Stubkjaer, K.E., "Multifunctional fiber-optic microwave links based on remote heterodyne detection," IEEE Trans. Microwave Theory Tech., vol. 46, no. 5, pp. 458-468, May 1998.

[34] Yao, J., "Microwave Photonics: Photonic Generation of Microwave and Millimeter-wave Signals" International Journal Of Microwave And Optical Technology, Vol.5 No.1 January 2010, p. 16-21.

[35] Ramos, R.T., Seeds, A.J., "Fast heterodyne optical phaselock loop using double quantum well laser diodes," Electron. Lett., vol. 28, no. 1, pp. 82–83, Jan. 1992.

[36] Komerath, N.M., and Komerath, P.P., "Implications of Inter-Satellite Power Beaming using a Space Power Grid". Proceedings of the IEEE/ AIAA Aerospace Conference, Big Sky, Montana, March 2011.

[37] Komerath, N.M., Chowdhary, C.G., "Retail Beamed Power for a Micro Renewable Energy Architecture: Survey", Proceedings of the ISED conference, Bhubhaneshwar, India, Dec. 2010.

New Electric Online Oil Condition Monitoring Sensor – an Innovation in Early Failure Detection of Industrial Gears

Manfred MAUNTZ
cmc Instruments GmbH; Hauptstraße 388
Eschborn; 65760, Germany

and

Ulrich KUIPERS
South Westphalia University of Applied Sciences, Haldener Straße 182
Hagen, 58095, Germany

and

Jürgen GEGNER
University of Siegen, Paul-Bonatz-Straße 9-11
Siegen, 57076, Germany

ABSTRACT

A new online diagnostics system for the continuous condition monitoring of lubricating oils in industrial gearboxes is presented. Characteristic features of emerging component damage, such as wear, contamination or chemical aging, are identified in an early stage. The OilQSens® sensor effectively controls the proper operation conditions of bearings and cogwheels in gears. Also, the condition of insulating oils in transformers can be monitored. The online diagnostics system measures components of the specific complex impedance of oils. For instance, metal abrasion due to wear debris, broken oil molecules, forming acids or oil soaps result in an increase of the electrical conductivity, which directly correlates with the degree of contamination in the oil. The dielectrical properties of the oils are particularly determined by the water content that becomes accessible via an additional accurate measurement of the dielectric constant. For additivated oils, statements on the degradation of additives can also be derived from changes in the dielectric constant. For an efficient machine utilization and targeted damage prevention, the new OilQSens® online condition monitoring sensor system allows for timely preventative maintenance on demand rather than in rigid inspection intervals. The determination of impurities or reduction in the quality of the oil and the quasi continuous evaluation of wear and chemical aging follow the holistic approach of a real-time monitoring of a change in the condition of the oil-machine system.
Once the oil condition monitoring sensors are installed on the plants, the measuring data can be displayed and evaluated elsewhere.
The measuring signals are transmitted to a web-based condition monitoring system via LAN, WLAN or serial interfaces of the sensor system.
Monitoring of the damage mechanisms during proper operation below the tolerance limits of the components enables specific preventive maintenance independent of rigid inspection intervals.

Keywords: conductivity, impedance, diagnostic, web-based monitoring, chemical aging, additive degradation

1. PRINCIPLES OF OILQSENS®

With the OilQSens® unit, components of the complex impedances *X* of oils, in particular the specific electrical conductivity κ and the relative permittivity ε_r as well as the oil temperature *T* are measured [1-3]. The values κ and ε_r are determined independently of each other. Fig. 1 shows the sensor with its triple plate design.

Fig. 1: Sensor with triple plate design.

Oils are electrical non-conductors. The electrical residual conductivity of pure oils lies in the range below 1 pS/m.

For comparison, the electrical conductivity of the electrical non-conductor distilled water is larger by six orders of magnitude.

Abrasive (metallic) wear, ions, broken oil molecules, acids, oil soaps, etc., cause an increase of the oil conductivity κ. It rises with increasing ion concentration and mobility. The electrical conductivity of almost all impurities is high compared to the extremely low corresponding property of original pure oils. A direct connection between the electrical conductivity and the degree of contamination of oils is found. An increase of the electrical conductivity of the oil in operation can thus be interpreted as increasing wear or contamination of the lubricant. The aging of the oil is also evident in the degradation of additives. The used additives reveal high conductivity compared to oil.

The consumption of the additives is reflected in a reduction of the electrical conductivity and permittivity of the oil. The gradient, i.e. the time derivative, of the conductivity or the dielectric constant progression respectively represents a measure of the additive degradation and consumption. The full additive degradation is indicated by the slope of zero (bathtub curve). Then the measurement signal increases further with increasing pollution, water entry, etc.

Ion mobility and thus electrical conductivity κ are dependent on the internal friction of the oil and therefore also on its temperature. The conductivity κ of the oil increases with temperature. The type of pollution and its temperature dependence cannot be assumed to be known. To improve the comparability of measurements, a self-learning adaptive temperature compensation algorithm is necessary. An integral alteration of the oil quality can then be assessed by the temperature compensated conductivity value, whereas the type of contamination is not determinable. The relative permittivity is measured with the same basic sensor

arrangement as used for the determination of the electrical conductivity.

2. TRIAL OF THE SENSOR PERFORMANCE ON A ROLLING BEARING TEST RIG

Various stress cycles are run and speeds and torques measured on a bearing and gear test rig, shown in Fig. 2.

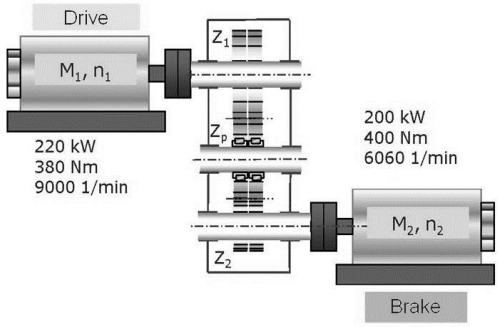


Fig. 2: Scheme of the used bearing and gear test rig.

In the case of an initially low stress, the electrical conductivity increases linearly with time. It is suggested that the low bearing wear in this range also increases with time.

During the necessary shutdown and run-up of the drive machine to 151 Nm, the conductivity is practically constant. After switching over to the higher load, the oil conductivity increases strongly. Here the effect of bearing run-in (shakedown) is shown as reduction in the conductivity increase. After switching off the test gear, the oil conductivity decreases markedly. This clearly underlines the influence of the additives. During the loading phases, more impurities per time unit are produced than can be bound by the additives. After shutting down the test rig, such oil contamination no longer occurs while the effect of additives still continues.

At the beginning of the stress, see Fig. 4, there is a run-in (shakedown) of the bearing. There it comes to small changes in the structures of individual components and a reduction in the quality of the machine. The stress will be increased to all components of the machine (like bearing) after the run-in and further more small stress at the bearing. The quality of the machine becomes significantly worse. Just in the end, shortly before the switch off, the connected measurement system of vibration analysis responds, long after the damage started.

Fig. 3 shows the inner ring of the failed planetary gear bearing. It reveals massive spalling damage on the raceway at the end of the rig test.

Fig. 3: Spalled inner ring raceway of the tested cylindrical roller bearing.

The quality of the test machine during the entire course of the experiment can be described qualitatively. Fig. 4 present this scheme.

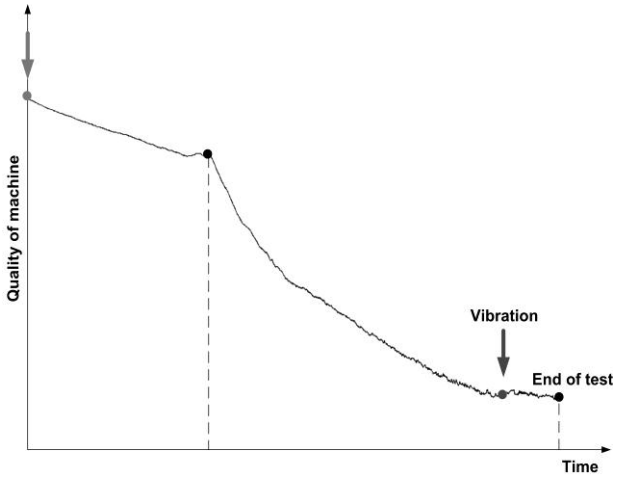


Fig. 4: Quality of the machine during the rig test.

Moreover, Fig. 5 displays the temperature-compensated oil conductivity as a function of the running time during the performed gearbox rig test.

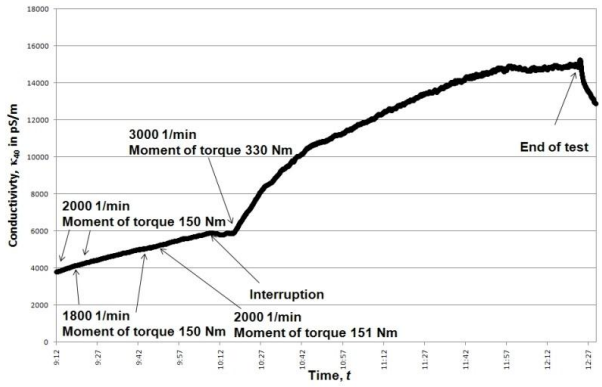


Fig. 5: Course of temperature compensated conductivity of the oil.

Figure 6 displays the gradient of electrical conductivity.

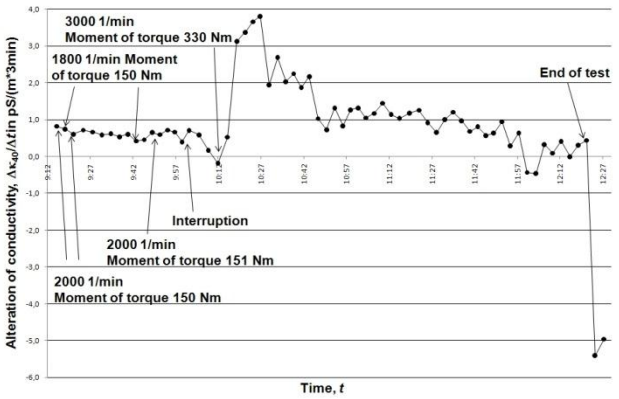


Fig. 6: Gradient of the electrical conductivity.

When starting up with a speed frequency of 2000 min^{-1} and a torque of 150 Nm, a relatively constant alteration of the conductivity from 0.6 to 0.8 pS/(m·3 min) occurs. After increasing the load to 330 Nm at 3000 min^{-1}, the change in conductivity raises up to 3.8 pS/(m·3 min). After the intermediate load increase, the effect on the change of the oil conductivity appears stronger. This may be attributed to the time-dependent formation of impurities and changes in bearing stressing as can be expected during the development of spalling. The real-time monitoring of the state change of the oil-machine system allows a specific control of the machine in terms of its wear limit. Improper loading conditions in inadmissible wear ranges can thus be avoided, which results in a more efficient machine utilization.

3. APPROACH FOR CONDITION MONITORING OF ADDITIVATED LUBRICATING OILS

Fig. 7 schematically shows the temperature compensated time curve of the permittivity of additivated oil continuously contaminated by the addition of wear debris, water or oil acids from chemical aging. Once the additives are consumed, the vanishing shielding effect results in a characteristic re-increase.

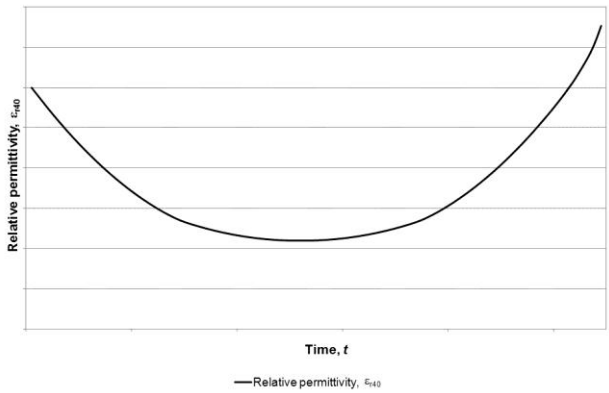


Fig. 7: Temperature compensated permittivity.

The most commercially available particle counters can only detect particles as small as 4 µm. In a very early stage of wear of bearings, gears, hydraulic cylinders, etc., however, particularly smaller particles are produced. A preventive maintenance lowing, rather than rigid inspection intervals, therefore requires recognition of even the smallest particles. These particles are far more common in the oils of functioning machines than larger ones. Oil aging can be involved in the failure, for instance, of rolling bearings [4].

4. WEB-BASED DECENTRALIZED LUBRICANT QUALITY MONITORING SYSTEM

The integration into a suitable communication structure and the realization of an online monitoring system offers an interesting practice-oriented utilization of the oil sensor system. This is briefly discussed below.

Preferred areas of application of the sensor system are energy production and automated technical plants that are operated locally, like e.g. wind turbines, transformer stations, generators, hydraulic systems or gearboxes. Plant employers are interested in continuous automated in vivo examination of the oil quality rather than interrupting the operation for regular sampling. Online oil status monitoring significantly improves the economic and ecological efficiency by increasing operating safety, reducing down times or adjusting oil change intervals to actual requirements. Once the oil condition monitoring sensors are installed on the plants, the measuring data can be displayed and evaluated elsewhere. A flexible decentralized monitoring system also enables the analysis of measuring signals and monitoring of the plants by external providers. A user-orientated service ensuring the quantitative evaluation of changes in the oil-machine system, including the recommendation of resulting preventive maintenance measures, relieves plant operators, increases reliability and saves costs.

In a web-based decentralized online oil condition monitoring system, the sensor signals are preferably transferred through the Internet to a database server and recorded on an HTML page as user interface.
Fig. 8 shows the displayed measured data.

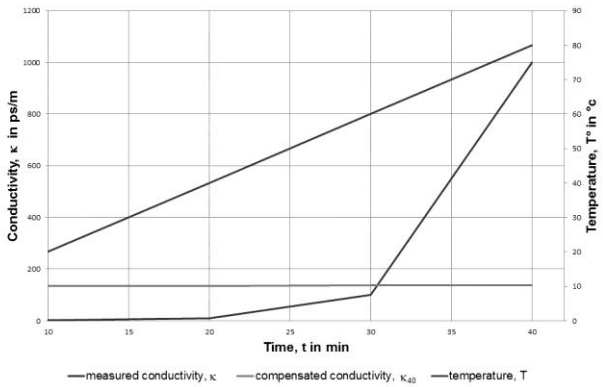


Fig. 8: The displayed measured data.

Following authentication, a simple web browser permits access via the wired or wireless LAN. In case of alarm signals, an immediate automated generation of warning messages, for instance by e-mail or SMS, is possible from any computer with Internet connection. Fig. 9 shows the OilQSens® sensor system with communication module [5].

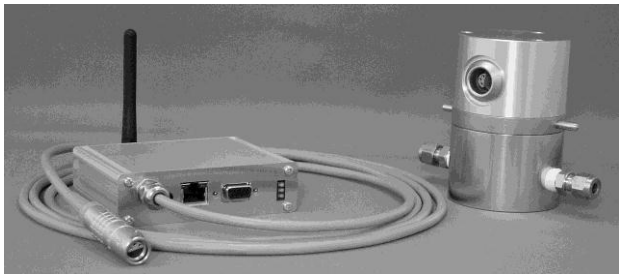


Fig. 9: OilQSens® sensor system with communication module.

5. SUMMARY

The online diagnostics system measures components of the specific complex impedance of oils. For instance, metal abrasion due to bearing wear at the tribological contact, broken oil molecules, acids or oil soap cause an increase in electrical conductivity that directly correlates with the degree of pollution of the oil. The dielectrical properties of the oils are especially determined by the water content, which, in the case of products that are not enriched with additives, becomes accessible by an additional accurate measurement of the dielectric constant. In the case of oils enriched with additives, statements on the degradation of additives can also be deduced from recorded changes in the dielectric constant.

Indication of wear is measured as an integral factor of, e.g., the degree of pollution, oil aging and acidification, water content and the decomposition state of additives, abrasion of the bearings.

It constitutes an informative standard of lubricant and material stress as well as the wear of the bearings and gears for the online operative monitoring of components of machines. Additional loading, for instance, by boundary lubrication in rolling-sliding contact (rolling bearings, gears, cams, etc.) causes the oil to age faster. The oil experiences incipient resinification in roller bearing tests and significant acidification is proven by infrared spectroscopy.

For an efficient machine utilization and targeted damage prevention, the new OilQSens® online condition monitoring system offers the prospect to carry out timely preventative maintenance on demand rather than in rigid inspection intervals. The determination of impurities or reduction in the quality of the lubricants and the quasi continuous evaluation of the bearing and gear wear and oil aging meet the holistic approach of a real-time monitoring of a change in the condition of the oil-machine system.

The measuring signals can be transmitted to a web-based condition monitoring system via LAN, WLAN or serial interfaces of the sensor system. The monitoring of the tribological wear mechanisms during proper operation below the tolerance limits of the components then allows preventive, condition-oriented maintenance to be carried out, if necessary, long before regular overhauling, thus reducing outages caused by wear while simultaneously increasing the overall lifetime of the oil-machine system.

On a bearing and gear rig test, various load cycles are run and the functionality of the introduced electric online condition monitoring sensor system is tested successfully. The evaluation of the experiment is presented.

REFERENCES

[1] J. Gegner, U. Kuipers, M. Mauntz, **Ölsensorsystem zur Echtzeit-Zustandsüberwachung von technischen Anlagen und Maschinen.** Technisches Messen 77 (2010) 283-292

[2] U. Kuipers, M. Mauntz, **Ölsensorsystem – Sensorsystem zur Messung von Komponenten der komplexen Impedanz elektrisch gering leitender und nichtleitender Fluide, dessen Realisierung und Anwendung.** Patentanmeldung Nr. 10 2008 047 366.9, Anmelder: cmc Instruments GmbH, Deutsches Patentamt, München, Anmeldetag: 15. 9. 2008

[3] U. Kuipers, M. Mauntz, **Verfahren, Schaltungsanordnung, Sensor zur Messung physikalischer Größen in Fluiden sowie deren Verwendung.** Europäische Patentanmeldung Nr. EP 09000244, Europäisches Patentamt, München, 9. 1. 2009

[4] J. Gegner, W. Nierlich, **Mechanical and Tribochemical Mechanisms of Mixed Friction Induced Surface Failures of Rolling Bearings and Modeling of Competing Shear and Tensile Stress Controlled Damage Initiation.** T+S Tribologie und Schmierungstechnik 58 (2011) 10-21

[5] J. Gegner, U. Kuipers, M. Mauntz, **Online Condition Monitoring System OilQSens® for continuous monitoring and identification of tribological wear mechanisms in the oil of bearings and gears.** Sensor+Test Nürnberg 2011, 7. - 9. Juni 2011, 08.06.2011

Construction Safety Visualization

Pramen SHRESTHA, Evangelos YFANTIS, & Kishor SHRESTHA
Civil and Environmental Engineering Department & School of Computer Science
University of Nevada, Las Vegas (UNLV)
Las Vegas, Nevada 89154, U.S.A.

ABSTRACT

Throughout the history of the construction industry, many fatalities and injuries have occurred in construction sites. One of the major causes of accidents is unsafe site conditions, which basically is due to inadequate supervision. To improve upon the traditional supervision approach, this study proposes a construction safety visualization approach. In this research paper, we provide a computer vision algorithm to detect in real time if any safety violations occur caused by one or more people in the site not wearing their hard hats. Our algorithm is an extension of our previous work in computer vision, and consists of a detection algorithm that is relatively fast and yet has an accurate edge. This is followed by a video object plane algorithm for each person, which is followed by an algorithm for detecting if a person does or does not wear a hard hat.

Keywords: Image processing, Hard Hats, Edge Detection Algorithm, Wireless Digital Camera, Local Area Network.

1. INTRODUCTION

There has been a strong myth that accidents are inevitable during construction projects. Nowadays, this notion is fading out. An innovative idea is emerging that zero injuries at construction sites are possible. As injuries and accidents are quite costly as well as hurt the morale of the workers, contractors have been trying to create a zero-injury culture in the construction industry; however, much effort is still necessary to avoid construction accidents. There are several categories of injuries at a construction site: falling injury, electrical injury, and injuries due to being struck by falling objects or flying objects, to name a few. In 2009, 4551 people died while working in construction sites. Of these, about 728 people (16%) died due to being struck by the objects and equipment [4]. Therefore, it is necessary that construction workers use hard hats on construction sites for head protection. Construction safety training teaches construction workers to use hard hats in their job sites. Also, the contractor does not allow anybody on the construction site without wearing hard hats. However, even though the use of hard hats by workers is mandatory, there are numerous occasions where workers had fatal accidents due to not wearing hard hats. Therefore, the enforcement of hard hat use in the construction sites is getting stricter. For example, the safety engineer or construction supervisor has to monitor whether their workers are using hard hats or not. Due to pressure in completing the project on time, and also due to complex work schedules, it is becoming difficult for supervisors to monitor every worker to check whether they are using safety equipment at the sites. To overcome this deficiency, if a visualization algorithm can be developed so that the real time image received from the sites can be analyzed to check whether the workers are using hard hats, it will assist supervisors to save the lives of their workers. Thus, the major focus of this research is to explore a visualization method that could determine whether workers are wearing hard hats at the construction sites.

To improve construction safety, Occupational Safety and Health Administration (OSHA) provides safety training to construction workers and also prepares construction safety standards. Due to the intervention by OSHA, fatalities at construction sites are decreasing. However, construction managers still are worried because construction fatalities and injuries have not reduced significantly enough. OSHA has been educating construction workers to use personal protective equipment (PPE), such as hard hats, safety shoes, harnesses, goggles, face shields, reflective clothing, filter masks, and ear plugs. The use of these devices is very important to protect constructions workers from injury and fatality. It is necessary to monitor in real time the construction site in order to make sure that the workers are using PPEs.

The visualization technique, which uses real-time images coupled with computer algorithms, is a useful tool to monitor the construction workers and warn supervisors if they are not using hard hats. In this technique, cameras are installed in the construction sites and the real-time images are transferred to the computers by means of wireless technology. The images of the construction site are continuously displayed on an office computer. From the real-time images, the algorithm will detect whether

the construction workers are using hard hats. Once the algorithm identifies a worker working without hard hat, it automatically dispatches a warning message to the safety officer. Ultimately, the site supervisor that is responsible will be informed so that the safety problem is corrected prior to an accident occurring.

Computer vision is a large area that applies to many areas, such as medicine, military, automation, transportation and construction management. The use of computer vision could help increase the productivity, decrease loss of property, and improve safety at the construction site. In this research, our main concern is real-time automatic detection on whether or not people are wearing their hard hats. If one or more people are not wearing their hard hats, that will constitute a safety violation. Such a safety violation will be recorded in a safety violation data base, along with the time and duration of the violation. Furthermore, this violation will issue an alarm event, which will be shown on the alarm events monitor for the onsite manager to see and take corrective action. Alarm events such as this one could also be transmitted to cell phones of specified people or else their office computers to make them aware of violation. The system that was developed in this study consists of cameras using charge couple devices to capture high-quality, uncompressed analog video at NTSC resolution. In the back of each camera is a hardware electronic card specifically designed to capture the camera video and convert it to digital; the card also processes the video by applying a detection algorithm and compresses the video, all in real time. The electronic board that was developed in this study detects hard hat violations and transmits them to the file server; a special flag indicates the type of violation detected, in this case, a hard hat violation. This violation, which is stored on a special data base in the file server, also triggers a software event alarm. The images associated with this alarm event are automatically shown in the special event monitor. If a network of cameras is used with proper overlap, a stereo vision can provide the exact location in the construction site, the name of the person(s) causing the violation, the duration of the violation, and the history of each person relating to safety violations. This camera network system also can be used in other applications to deter loss of property, detect fraudulent accidents, and avoid accidents; this can be done by analyzing activities real time and issuing alarm events on activities that have the potential to cause accident.

This current study is an extension of previous, related work in the areas of computer vision, robot vision, image compression, pattern recognition, internet transmission, network communication, and image processing [1, 3, 6, 7].

2. BACKGROUND

Visualization techniques have been used in construction planning and operations. A study was conducted to demonstrate use of a simulation-driven, dynamic 3D visualization process in a multi-storey structural steel erection operation [8]. The authors used visualization techniques that employed dynamic operations to depict the continuously evolving multi-storey structural steel facility. Four-dimensional computer-aided design (CAD) visualizations only show the evaluation of the construction product and can be linked to project schedules. However, dynamic operation visualization can show the interactions between various resources, including machines, materials, and temporary structures. The authors showed how this process can help the contractors build the project more efficiently and effectively.

In 2009, the Construction Industry Institute (CII) conducted a study to use pro-active real time safety technology on the equipment and worker to warn them about possible accidents [2]. They implanted in heavy equipment very-high-frequency active radio frequency (RF) technology, consisting of an in-cab device and a personal device. The personal protection unit (PPU) used by construction workers consisted of a chip, a battery, and an alarm. When the workers are in the proximity of the heavy equipment, the alarm is set off in the equipment as well as on the workers' PPU. The field tests demonstrated that by implementing this technology, various benefits can be achieved, for instance, providing real-time pro-active alerts to workers and operators and also monitoring the locations of workers, equipment, and materials. Moreover, this study included a cost-benefit analysis to show that it is economically viability to use real-time pro-active technology in the construction sites.

A study was conducted by Michal Irani and P. Anandan [5], in which the detection of moving objects was carried out. In the past, many studies of this kind were done by using 2D algorithms; this study used a unified approach to handle moving object detection in both 2D and 3D scenes, with a strategy that gracefully bridged the gap between those two extremes. According to the author, 3D algorithms work better when significant depth variations are present in the scene and the camera is translating. This approach is based on a stratification of the moving object-detection problem into scenarios that gradually increase in their complexity.

3. SEGEMENTATION AND REGION OF INTEREST DETECTION

The metric used in this research is a non-Euclidian metric. The mathematical space we operate on is a Banach space. In this Banach space, we define a

probability metric later on. Let I(x,y) be the intensity of the image at position (x,y); then, an estimate of the second partial derivative with respect to x is:

$$\frac{\partial^2 I(x,y)}{\partial x^2} = \frac{I(x-1,y)-2I(x,y)+I(x+1,y)}{2} \quad (1)$$

and an estimate of the

$$\frac{\partial^2 I(x,y)}{\partial y^2} = \frac{I(x,y-1)-2I(x,y)+I(x,y+1)}{2} \quad (2)$$

The Laplacian, or divergence, of the gradient at the point (x,y) of the gray scale image is:

$$\Delta \mathrm{I}(\mathrm{x},\mathrm{y}) = \nabla^2 I(x,y) = \frac{\partial^2 I(X,y)}{\partial x^2} + \frac{\partial^2 I(x,y)}{\partial y^2} \quad (3)$$

From equations (1) an (2) we obtain an estimate of the Laplacian of the gray scale image at pixel position (x,y):

$$\Delta I(x,y) = \nabla^2\, I(x,y) =$$

$$\frac{I(x-1,y)+I(x+1,y)+I(x,y-1)+I(x,y+1)-4I(x,y)}{2} \quad (4)$$

The values of the intensity all are integers in between 0 to 256. Multiplication by 4 can be obtained by shifting the integer two times to the left. Division by 2 is obtained by first adding 1 to the numerator if the numerator is positive or subtracting 1 from the numerator if negative, and then shifting the numerator to the right one. The estimated Laplacian for any image could be negative or positive, with the majority of the values being equal to zero and is symmetric about zero. The probability density function of the Laplacian is:

$$f(x) = \frac{1}{\sqrt{2}\sigma} e^{-\frac{\sqrt{2}|x|}{\sigma}} \qquad -\infty < x < \infty \quad (5)$$

The standard deviation σ depends on the quality of the camera, the light intensity of the scene, and the number of edges as well as the type of edges. For example, edges both of metallic objects and steel objects reflect light differently than edges of non-steel material. The edges of an image represent a relatively small percentage of the pixels of the image; those points are part of the tails of the probability density function of Equation (5). Our segmentation algorithm consists of finding the edges by first using the above theory.

Edge Detection Algorithm

1. Compute the Luma component of the image.
2. For every Luma component pixel, compute the second order partial derivative with respect to x. using Equation (1).
3. For every Luma component pixel, compute the second order partial derivative with respect to y, using Equation (2).
4. For every Luma component pixel, compute the Laplacian at position (x,y), using Equation (3).
5. Compute the histogram of the values obtained in Equation (4).
6. All the values for which the area of the histogram to the right is less than 2.5% are edges.
7. All the values for which the area of the histogram to the right is less than 10% are possible edges.
8. If any of the neighbors of a possible edge is an edge then the possible edge is an edge.

In order to segment an object or subject of interest, first the edge detection is used to outline the object, and then the necessary rules are provided in order to separate the outline of the object from other outlines. In this case, the outline of a hard hat depends on the angle of the camera with respect to the hard hat.

Segmentation Algorithm

1. Use the Edge Detection Algorithm described above.
2. Divide the scene into Video Object Planes (VOPs).
3. Separate the VOPs according to the people in the scene.
4. Focus on the part of the VOP of a person between the base of the neck and the upper end.
5. The sub-video object plane of the hard hat consists of two orthogonal semicircles.
6. The cord line joining the ends of the top semicircle forms an angle with the x-axis of the screen. This angle θ can be determined by taking the dot product of the cord line row vector with the row vector (1, 0).
7. If the above normalized dot product is one, then the cord is parallel to the screen coordinate x-axis and the person is standing straight up; in this case, there is a possibility of the head tilting towards the left or right shoulder. Notice that with more than one camera, this possibility can be resolved.
8. If the dot product is a number different than zero and between -1 and 1, then the hard hat can be rotated by -arccosθ about the x axis into a normalized position.
9. Logos or initials painted in the hard hats are segmented and recognizable during the segmentation.
10. If more than one camera have the same object in view, then stereo vision of the hard hat is possible. In this case, straps, logos, the pitch, role, and yaw of the hard hat are computable in real time.
11. If the cameras are properly calibrated and the distance between cameras is known, then the distance of the various objects with respect to an origin can be computed in real time.

4. APPLICATIONS

Figure 1 shows the part of a construction scene that was obtained by one camera. The video object plane of the person depicted in the scene was obtained by first applying the edge detection algorithm, then applying the segmentation algorithm, part of which is the video object plane detection for a person and the sub video object segmentation for the hard hat. The two characteristic semicircles of the hard hat can be seen, namely the upper semicircle whose cord forms an angle with the x-axis, and the slightly deformed horizontal semicircle. The relative clarity of the safety jacket with the characteristic stripes as well as the clear definition of the right ear, mouth, nose, right eye, right arm, and right hand all can be noticed in the picture. All of these are part of the features included in the automatic, or computerized, hard hat recognition algorithm. Although one camera connected to a file server, which applies the pattern recognition algorithm to detect each person in the view of the camera and subsequently decides in real time if each person in the view of the camera has the hard hat on or not. If two or more cameras are used to view a person simultaneously from two different angles, and if the distance between any two cameras is known and fixed, then by applying camera calibration prior to using the cameras for detecting if the people in the construction scene wear their hard hats or not, and if we assign an arbitrary coordinate system with an arbitrary coordinate system origin, then distances can be resolved of people from the origin of the coordinate system. Stereo video object planes of people also can be produced.

Figure 2 was produced from a colored picture of a construction scene that included several workers, all wearing their hard hat except one. First, edge detection algorithm was applied and then the video object plane was applied to separate each person in the camera view; finally, the hard hat algorithm detected no hard hat for the person shown in Figure 2. Notice that the upper semicircle is similar to that of the hard hat. It is different from the hard hat semicircle. Those differences are that it is larger than a semicircle; it is not as smooth as a hard hat semicircle; it does not provide the discontinuity with the rest of the head and neck that the hard hat has; and it does not have the lower profile of a hard hat.

Fig. 1. The video object plane of person and the sub-video object plane of the hard hat.

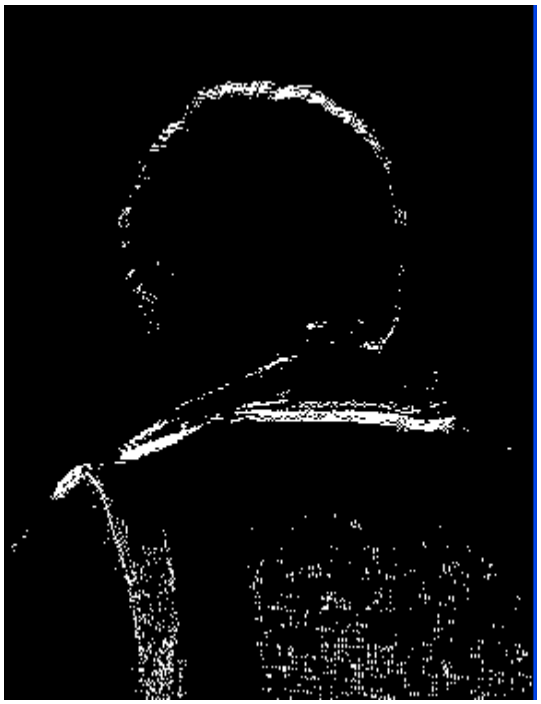


Fig. 2. A worker with no hard hat.

Figure 3 shows the upper semicircle, the base, and also the initials (UNLV) in the hard hat. Edge detection and image segmentation algorithms work which we are proposing work with all hard hats. Preliminary results show that the hard hats with safety stripes in the back, certain indentations in their design, and logos are easier to recognize as they provide additional information to our pattern recognition system. Camera systems are relatively inexpensive, they could be self powered, and easy to install. They could be connected to our on site and from there transmit wireless to offsite computers for automatic monitoring of productivity, and safety by automatically monitoring if everybody wears their hard hats and other protective gear. In case there is a violation it can issue a violation alarm at the computer screen and make the supervisors aware of the violation so that they can take corrective action.

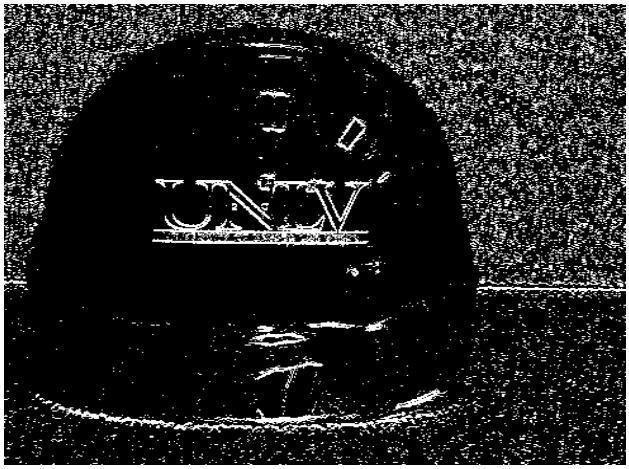


Fig. 3. Hard hat on the ground of the construction site

The system consists of a number of camera batteries that are powered by a combination of solar and other battery sources. Each camera is attached to card for video capture and video processing. The purpose of the card is to capture the video, analyze it for safety, compress the video, check for productivity, and transmit the video and the events wirelessly to a local file server. The file server is also powered by a combination of solar panels and lithium batteries. There are a number of safety events generated by the camera. The one pertaining to this research is the detection of one or more people in the construction site who are not wearing their hard hats. In such cases, the time of the event, along with event itself, is transmitted to a special safety event database in the server. Safety event violations are programmed to create an alarm event, which alerts the onsite construction manager by sending the event to that person's mobile phone. In the case of multiple cameras, the picture of the person causing the safety violation, along with their name, is transmitted to the construction manager as well as all other people that are set up in the system to receive this event violation.

This system easily can be set up to record productivity of people by recording automatically the total time per day that they actually work compared to the time engaging in non-work activities. Furthermore, the system could also monitor accidents and their causes, which might prevent compensation abuse due to fraudulent accidents.

5. CONCLUSIONS AND RECOMMENDATIONS

An algorithm for detecting hard hat safety violations was described in this research paper. This algorithm involves digital image processing, artificial intelligence, and stochastic pattern recognition. The implementation of the algorithm involves electronic design, electronic drivers, embedded systems, file servers, internet software, network software, and data base management. The purpose of the system is to make construction managers, and others who need to know any safety violations that are a result of one or more people on the site not wearing their hard hats. This system can be used for other safety violations as well as for productivity detection.

A computer system and algorithm developed through this research analyzes real-time images of the construction site in order to detect workers working without hard hats. A demonstration will be done at UNLV's Construction Management Laboratory for the validation of the system before using in the real sites. This technique was developed for the detection of workers wearing hard hats with an aim of significantly reducing injuries at construction sites.

In the future, the authors would like to expand this research for detection of all personal protective equipment (PPEs) as well as heavy equipment used at construction sites so that workers can be safer and more productive.

6. REFERENCES

[1] A. Ritchie, J. Conradi, A. Prevot, and E.A. Yfantis "Robot Vision and Video Transmission," **Transactions on Communication**, ISSN 1109-2742, Issue 8, Vol. 9, August 2010, pp. 515-524.

[2] Construction Industry Institute, **Real-time Pro-active Safety in Construction,** Research Summary 269-1, Version 1.1, August 2010.

[3] E.A. Yfantis, M. Y. Au & G. Miel, "Efficient Image Compression Algorithm for Computer Animated Images," **Journal of Electronic Imaging**, Vol. 1, No. 4, 1992, pp. 381-388.

[4] Occupation Safety and Health Administration, **2009 Census of Fatal Occupational Injuries.** United States Department of Labor, Bureau of Labor Statistics, March, 2011.
<http://www.bls.gov/iif/oshcfoi1.htm#2009>

[5] M. Irani and P. Anandan, "A Unified Approach to Moving Object Detection in 2D and 3D Scenes," **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol. 20, No. 6, 1998, pp. 577-589.

[6] Yfantis, E. A. "A new Quadratic and Biquadratic Algorithm for Curve and Surface Estimation," **Journal of Computer Aided Geometric Design**, Vol. 10, 1993, pp. 509-520.

[7] Yfantis, E. A., **Visually Lossless Still Image Compression for CMYK**, CMY and Postscript Formats, Patent no. US 6,633,679 BI 10-14-2003.
V. R. Kamat, M. Golparvar-Fard, J.C. Martinez, F. Pena-Mora, M. Fischer, and S. Savarese "Research in Visualization Techniques for Field Construction**," Journal of Construction Engineering and Management**, ASCE, 2010.

AUTHORS INDEX
Volume I