

ISSN: 1823-3287



RPCES

Johor Bahru, July 26-27, 2006

Universiti Teknologi Malaysia

REGIONAL POSTGRADUATE CONFERENCE ON
ENGINEERING & SCIENCE



School of Graduate Studies (SPS)
Indonesian Student Association (PPI)
Universiti Teknologi Malaysia



ORGANIZING COMMITTEE

PATRON

Datuk Prof. Ir. Dr. Mohd Zulkifli Tan Sri Mohd Ghazali
(Vice Chancellor, Universiti Teknologi Malaysia)

ADVISORY BOARD

Prof. Dr. Rahmalan Ahamad
(Dean SPS, Universiti Teknologi Malaysia)

Prof. Dr. Ahmad Fauzi Ismail
(Universiti Teknologi Malaysia, Malaysia)

Assoc. Prof. Dr. Nordin Yahaya
(Universiti Teknologi Malaysia, Malaysia)

Assoc. Prof. Dr. Waluyo Adi Siswanto
(Muhammadiyah University of Surakarta, Indonesia)

Assoc. Prof. Dr. Hishamuddin Jamaluddin
(Universiti Teknologi Malaysia, Malaysia)

TECHNICAL COMMITTEE

Prof. Dr. Ahmad Fauzi Ismail
(Universiti Teknologi Malaysia, Malaysia)

Assoc. Prof. Dr. Nordin Yahaya
(Universiti Teknologi Malaysia, Malaysia)

Assoc. Prof. Dr. Nor Abidah Mohd Omar
(Universiti Teknologi Malaysia, Malaysia)

Assoc. Prof. Dr. Waluyo Adi Siswanto
(Muhammadiyah University of Surakarta, Indonesia)

Assoc. Prof. Dr. Abdullah
(Diponegoro University, Indonesia)

Dr. Hadi Nur
(Universiti Teknologi Malaysia, Malaysia)

Dr. Nurly Gofar
(Universiti Teknologi Malaysia, Malaysia)

WORKING COMMITTEE

Chairman	Tutuk Djoko Kusworo
Deputy Chairman	Reza Firsandaya Malik Abdullah Saand
Secretariat	Kusmiyati Norly Mazlan Ammar Yassin Al-Mukrary Denny Kurniawan
Treasurer	Jarot Setyowiyoto
Paper/Proceeding	Amrifan Saladin Mohruni Megat Aman Zahiri Ngurah Made Dharma Putra Izham bin Mohammad Yusoff
Conference Program	Ardiyansyah Syahrom Djurdjani
Information Technology	Istadi Dahliyusmanto
Publication	Agung Murti Nugroho
Logistics/Accomodation/ Transportation	Prijono Nugroho M. Ni'am Faiqun
Souvenirs	Ahmad Nasirudin
Food & Beverages	Azriyenni Salmiati

Table of Contents

A Construction and Build Environment

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
A1	Need of Sustainability in Construction Industry of Pakistan	<u>S.Mubin</u>	sajjadmubin@uet.edu.pk	Faculty of Civil Engineering University of Engineering and Technology, Lahore, Pakistan	1
A2	Semi Continuous Composite Connection Using Trapezoid Web Profiled (TWP) Steel Beam Section	Anis Saggaff, Mahmood Md Tahir, Shahrin Muhammad, Sariffuddin bin Sa'ad, Arizu Sulaiman .	anissaggaf@yahoo.com	Civil Engineering Department, Faculty of Engineering, Sriwijaya University, Palembang, Indonesia	7
A4	Effect of Aggregate Gradation to workability of hot asphalt mixture	Sri Widodo , Ika Setianingsih	sri_widodo@ums.ac.id	Department of Civil Engineering, Muhammadiyah University of Surakarta Indonesia Tel: +62-271-717417, Fax +62-271-715448	15
A5	The Effect of Road Section Characteristics on Safety Belt Use Rate in Jakarta	<u>Leksmono Suryo Putranto</u> , Budi Fitrianto, Achmad Khusaini	lexy@tarumanagara.ac.id	Department of Civil Engineering Tarumanagara University, Jl. Let. Jen. S. Parman No. 1, Jakarta 11440, Indonesia Tel: +62-21-5672548, Fax: +62-21-5663277,	21
A6	Laboratory Evaluation of Horizontal Coefficient of Consolidation of Fibrous Peat Soil	<u>Yulindasari Sutejo</u> , Wong Leong Sing , Nurly Gofar	indatejo@yahoo.com	Department of Geotechnics and Transportation, Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor	27
A9	Influence of Natural Organic Matter Characteristics on Fouling and Rejection Mechanisms of Immersed Ultrafiltration Membrane filtration. A Case Study of Potable Water Production	A.W. Zularisam, A.F. Ismail, M.R. Salim	afauzi@utm.my	Environmental Engineering Focus Group, FKA, Kolej Universiti Kejuruteraan & TeknologiMalaysia, KUKTEM, Gambang, Pahang, Malaysia	33
A10	The Effect of Wall Rigidity and Deformation on Lateral Earth Pressure Distribution	<u>Lee Min Lee</u> , Lo Seng Chew, Nurly Gofar	amuellee@yahoo.com	Faculty of Civil Engineering Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +607-5532517, +6012- 8808979	39
A11	Effect of Binder Sand Ratio to Basic Properties of Epoxy Grout	S. K. Lim, F. Zakaria, M. W. Hussin, Z. Abd. Hamid, R. Muhamad	siongang@hotmail.com	Faculty of Civil Engineering Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-7-5532158, Fax: +60-7-5531967,	45
A13	The Effect Of Palm Oil Fuel Ash Content And Water Dry Mix Ratio Towards The Density And Compressive Strength Of POFA Cement Based Aerated Concrete	K. Abdullah, M. W. Hussin, F. Zakaria, Z. Abd. Hamid, R. Muhamad	khairunisa1978@yahoo.com	Faculty of Civil Engineering Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-7-5532158, Fax: +60-7-5531967	51
A15	Separation of Suspended Particles from Wastewater using Electrocoagulation Technology	<u>Moh Faiqun Ni'am</u> , Fadil Othman , Johan Sohaili , Zulfa Fauzia	dfadilos@utm.my	Environmental Department., Faculty of Civil Engineering Universiti Teknologi Malaysia, Johor, Malaysia	59
A16	Integration of Constructability Principles into the Design Process	Woon, K. S., and	wkaisiong@yahoo.co.uk	Faculty of Civil Engineering Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor,	63

A Construction and Build Environment

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
		Rosli, M. Z.		Malaysia	
A17	The Effect of Joint Width between Blocks in Concrete Block Pavement	Rachmat Mudiyo, Hasanan Md. Nor and Ling Tung Chai	rachmatmudi@yahoo.com	Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-7-5531500, Fax: +60-7-5566157,	69
A18	An Overview of Construction Accidents in Malaysia	Abdul Rahim Abdul Hamid, Muhd Zaimi Abd Majid, Bachan Singh	rahimfka@yahoo.com	Faculty of Civil Engineering Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-7-5531689, Fax: +60-7-5566157	77
A20	Predicting the Performance of Design-Bid-Build Projects: A Neural-Network Based Approach	Caren C. L. Tan, Rosli M. Z,	tan_caren@hotmail.com	Faculty of Civil Engineering Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia	83
A21	The Effect of Cement and Water Cement Ratio on Concrete Paving Block	Tung-Chai Ling , Hasanan Md Nor , Rachmat Mudiyo	tcling611@yahoo.com	Department of Geotechnics and Transportation Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-7-5531580, Fax: +60-7-5566157	89
A22	Development of Sections Properties Tables for Cruciform Column Using Universal Beam Sections	M.M.Tahir, P.N. Shek, C.S. Tan, S. Anis	mahmoodmdtahir@yahoo.com	Steel Technology Centre, Faculty of Civil Engineering Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-7-5531616, Fax: +60-7-5576841	93

B Information and Communication Technology

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
B3	Modeling and Appointment Scheduling for Vehicle Inspection Center	Nuzulha Khilwani Ibrahim, Habibollah Haron, Nurulhuda Firdaus Mohd Azm	nuzulha@hotmail.com	Faculty of Computer Science and Information System Universiti Teknologi Malaysia 81310 UTM Skudai, Johor	101
B4	Features Selection for Probabilistic Model of Response Generation in Dialogue Systems	<u>Aida Mustapha</u> , Md. Nasir Sulaiman, Ramlan Mahmod, Hasan Selamat	{aida, nasir, ramlan, hasan}@fsktm.upm.edu.my	Faculty of Computer Science and Information Technology Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia Tel: +60-3-89466554, Fax: +60-3-89466577.	107
B10	Document Image Recognition Of Javanese Manuscripts (Case Study of Javanese Manuscripts: Menak Sorangan I and Panji Sekar)	Anastasia Rita Widiarti , Agus Harjoko	rita@staff.usd.ac.id	Faculty of Mathematics and Natural Sciences Sanata Dharma University, Paingan Maguwoharjo Depok Sleman, Yogyakarta 55002, Indonesia Tel: +60-0274-883968, Fax: +60-0274-886529	117
B11	Palmpriint Verification Based on Fractal Codes and Fractal Dimensions	Darma Putra, I Ketut Gede, Adhi Susanto, Agus Harjoko, Thomas Sri Widodo	eeit@telkom.ne , aharjoko@ugm.ac.id	Departement of Electrical Engineering, Faculty of Engineering, Udayana University, Bali – Indonesia	123
B14	Synthetic Aperture Radar and Optical-Sensor Data Fusion Based On A Trous Wavelet Transform	Wihartini and Aniat Murni	wihartini@bl.ac.id	Faculty of Engineering, Budi Luhur University Jl. Raya Ciledug, Jakarta, Indonesia Fax (62 21) 5853752,	129
B18	A Study on Model Adaption Technique for Improved Accuracy in Coded Speech.	Nachiappan and Abdul Manan Ahmad	nachi182001@hotmail.com	Faculty of Computer Science and Information Systems, University Technology Malaysia, Skudai, Malaysia. (Tel: +60(012)-6967860;	135
B21	Minimizing Handoff Latency in AMTree Protocol	<u>Ali M. Al-Sharafi</u> Mohamad Othman Shamala Subramaniam	a_alsharafi@hotmail.com	Faculty of Computer Science and Information Technology Universiti Putra Malaysia, Serdang 43400, Selangor D.E, Malaysia Tel: +603-89466565, Fax :+ 603-89466577 / +603-89466572	141
B23	A Study on Model Compensation Using Parallel Model Combination and Adaptation Technique for Improved Accuracy in Noisy Speech Recognition	Sang Sang Tan, Prof. Madya Abd. Manan Ahmad	ss_hello33@yahoo.com	Faculty of Computer Science & Information Systems Universiti Teknologi Malaysia 81310 UTM Skudai, Johor, Malaysia	145
B24	A Perspective of Manufacturing Features from Islamic Views	Setyawan Widyarto, Muhammad Saffie bin Md Yasan, Mohd Fahmi Mohamad Amran and Habibollah Haron	swidyarto@lycos.co.uk	Politeknik Negeri Lampung, Jalan Soekarno Hatta, Rajabasa, Bandar Lampung 35145, Indonesia Tel. 62-721-703995, Fax. 62-721-787309,	151
B25	A Physical Force Based Microscopic Crowd Model Simulation	Setyawan Widyarto, Aboamama Atahar, Muhammad Shafie Abdul Latif ²	swidyarto@lycos.co.uk	Politeknik Negeri Lampung, Jalan Soekarno Hatta, Rajabasa, Bandar Lampung 35145, Indonesia Tel. 62-721-703995, Fax. 62-721-787309	157
B26	Development of Spatio-temporal Query Language for Biodiversity Data	<i>Kamruzzaman, Mohd Taib Wahid, Harihodin Selamat</i>	akzaman@siswa.utm.my	Department of Information Systems, Faculty of Computer Science & Information Systems University Technology of Malaysia, K.B. 791 81310 Skudai, Johor, Malaysia	163

B Information and Communication Technology

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
				Tel: (607)-5532422, Fax: (607) 5565044	
B27	Local Positioning System of a Mobile Robot: A Practical Pre3spective	Iman H. Kartowisastro	imanhk@binus.ac.id	Computer Engineering Department, Bina Nusantara University, Jakarta, Indonesia. Tel : 62-21-534 5830, Fax : 62-21-530 0244	169
B28	Strategic Information Systems Planning (SISP) and Strategic Information Security Planning (SISecP) In Malaysian Government Agencies	Md Hafiz Selamat ^a , Mohd Adam Suhaimi ^b , Husnayati Hussin ^c	hafiz@fsksm.utm.my	Faculty of Computer Science and Information Systems Universiti Teknologi Malaysia, Skudai, Johor. Tel: +607-5532423, Fax: +607-5565044	175
B30	A Preliminarily Study on Embedded Platforms for Computer Vision Applications	Kenny Kuan Yew KHOO, Yong Haur TAY	kennykhooky@gmail.com	Universiti Tunku Abdul Rahman 13, Jalan 13/6. 46200, Petaling Jaya, Selangor	183
B31	Modeling Wind Speed Data in Surabaya and Sumenep, Indonesia Using Vector Autoregression (VAR)	Irhamah ^{1*} , Eling Anindita, Dwi Endah Kusrini	irhamah@yahoo.com	Department of Statistics, Institute of Technology Sepuluh Nopember, Kampus ITS Keputih, Sukolilo Surabaya, 60111, Indonesia	191

C Electrical and Electronics

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
C1	Variable Structure Control of Two-wheels inverted Pendulum Mobile Robot	Nawawi, S.W. Ahamad M.N., Osman, J.H.S.	sophan@fke.utm.my	Faculty of Electrical Engineering Universiti Teknologi Malaysia 81310 UTM Skudai, Johor	197
C2	Starfruit Grading Based on 2-Dimensional Color Map	M. M. Mokji, S.A.R. Abu Bakar	musa@fke.utm.my	Faculty of Electrical Engineering University of Technology MalaysiaJohor Bahru, Johor, Malaysia	203
C3	Overview of a Computer-based Stuttering Therapy	Ooi Chia Ai, J. Yunus	OoiChiaAi@yahoo.com	Department of Electronic Engineering Faculty of Electrical Engineering Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +6012-5844183,	207
C4	The Accuracy and Efficiency Issues of Decouple Approach for Harmonic Power Flow Calculation	Agus Ulinuha, Mohammad A. S. Masoum, Syed M. Islam	agus.ulínuha@postgrad.curtin.edu.au	Department of Electrical and Computer Engineering Curtin University of Technology, GPO Box U1987 Perth 684, Western Australia, Fax: +61-8-9266 2584	213
C5	Design of Multimode Interference Optical Splitter Based on BenzoCyclobutene (BCB 4024-40) Polymer	Mohd Haniff Ibrahim, Norazan Mohd Kassim, Abu Bakar Mohammad* & Mee-Koy Chin	hanif@fke.utm.my	*Photonic Research Group, Faculty of Electrical Engineering, Universiti Teknologi Malaysia. 81310 Skudai, Johor Malaysia.	219
C7	Analysis and Classification of Myocardial Infarction Tissue from Echocardiography Images Based on Texture Analysis	Nazori Agani, S.A.R. Abu-Bakar, and Sheikh Hussain Sheikh Salleh	nazori@bl.ac.id	¹ Department of Electrical Engineering Universitas Budi Luhur Jl. Raya Ciledug, Jakarta Selatan, Indonesia (12260) Telp. (021)5853753, Fax.(021)5853752	223
C9	Study of insulator to withstand and switching surges in conversion three to six phase transmission line computer simulation analysis	Muhammad 'Irfan Jambak ¹ , Hussein Ahmad ²	ijmuhammad2@siswa.utm.my	Faculty of Electrical Engineering Universiti Teknologi Malaysia UTM Skudai, 81310 Johor Bahru Malaysia	229
C10	Design and Modeling of On-Chip Planar Capacitor for Radio Frequency Application	Mariyatul Qibthiyah Mohd Noor and Mazlina Esa	mazlina@fke.utm.my	Microwave/RF and Antenna Research Group, Department of Radio Communication Engineering, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Darul Takzim, Malaysia. Tel: +60-7-5535262, Fax: +60-7-5566272	235
C11	Elimination of Mismatched Imbalance Vibration of Active Magnetic Bearing System Using Full Order Sliding Mode Controller	Husain A.R. ¹ , Ahmad M.N. ¹ and Mohd Yatim A.H. ¹	rashid@fke.utm.my	Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310, Skudai, Johor Bahru, Malaysia Tel: +60-7-5535894, Fax: +60-7-5544272	241
C12	Analytical and Empirical Study of Particle Swarm Optimization with a Sigmoid Decreasing Inertia Weight	Andi Adriansyah, Shamsudin H. M. Amin	aandi2@siswa.utm.my	Department of Electrical, Faculty of Industrial Engineering, Universitas Mercu Buana, Jl.Raya Meruya Selatan, Kembangan, Jakarta, 11650, Indonesia Phone: +6221-5840816, Fax : +6221-5840813	247

C Electrical and Electronics

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
C13	A True Random Number Generator for Crypto Embedded Systems	<u>Norashikin M. Thamrin</u> , Illiasaak Ahmad, Mohamed Khalil Hani	norashikin.mthamrin@gmail.com	VLSI-ECAD Research Laboratory, Microelectronic and Computer Engineering Department (MiCE) Faculty of Electrical Engineering Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia. Tel: +607-5535223, Fax: +607-5566272	253
C14	Accelerating Graph Algorithms with Priority Queue Processor	<u>Ch'ng Heng Sun</u> , Chew Eik Wee, Nasir Shaikh-Husin, Mohamed Khalil Hani	chnghengsun@yahoo.com.sg	VLSI-ECAD Research Laboratory (P04-Level 1), Microelectronic and Computer Engineering Department (MiCE), Faculty of Electrical Engineering (FKE), Universiti Teknologi Malaysia.	257
C15	Crypto Embedded System for Electronic Document	<u>Illiasaak Ahmad</u> , Norashikin M. Thamrin, Mohamed Khalil Hani	ilyasak_ahmad@yahoo.com	VLSI-ECAD Research Laboratory, Microelectronic and Computer Engineering Department (MiCE) Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia Tel: +60-7-5535223, Fax: +60-7-5566272,	263
C16	Nano-scale VLSI Clock Routing Module based on Useful-Skew Tree Algorithm	<u>Chew Eik Wee</u> , Ch'ng Heng Sun, Nasir Sheikh-Husin ² , Mohamed Khalil Hani	eikweechew@hotmail.com	VLSI-ECAD Research Laboratory (P04-Level 1), Microelectronic and Computer Engineering Department (MiCE), Faculty of Electrical Engineering (FKE), Universiti Teknologi Malaysia.	269
C17	Analysis of Throughput S-ALOHA CDMA with Differential MRC at Two Models Fading Channels	Hoga Saragih, Gunawan Wibisono and Eko Tjipto Rahardjo	hoga15876@yahoo.com	Department of Electrical Engineering, Faculty of Engineering, The University of Indonesia, Kampus Baru UI Depok, 16424, Jakarta, Indonesia	275
C18	Application of Fuzzy Logic in Active Vibration Control of Beam	Amirasyid D, Momoh Jimoh E.S	amirasyid@uniten.edu.my	Department of Mechanical Engineering, Faculty of Engineering Universiti Tenaga Nasional, 43009 Kajang Selangor Malaysia Tel: +60-31-89212020 Ext 6330, Fax: +60-3-89212116,	281

D Advanced Material and Materials Processing Technology

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
D1	Mode-I Interlaminar Fracture Toughness of Hand Lay-Up Plain-Weave Woven GFRP/Unsaturated Polyester Laminate	Mohd Aidy Faizal, Yeo Kiam Beng, Mohd Noh Dalimin	aidyfaizal_7513@yahoo.com	Center of Materials and Minerals (CMM) Universiti Malaysia Sabah, Beg Berkunci No. 2073, 88999 Kota Kinabalu, Sabah, Malaysia Tel. 088-320000 ext: 3543	287
D2	Fabrication of Polymer Light Emitting Diodes with ITO/PVK:TPP/Alq ₃ /Al Structure	C.C. Yap, M. Yahaya and M.M. Salleh	chichin83@yahoo.co.uk	Faculty of Science and Technol Institute of Microengineering and Nanoelectronics (IMEN) Universiti Kebangsaan Malaysia 43600 Bangi, Selangor, Malaysia Tel:+60-3-89213560, Fax:+60-3-89213777,	293
D3	Identifying Meshing Characteristics To Model Crack Tip Region For Linear Fracture Mechanics Application	E.H. Lim*, K. B. Yeo, Mohamed Harimi	ehlim@pasca.ums.com.my or stan_hel@hotmail.com	Centre of Materials and Minerals Universiti Malaysia Sabah, Locked Bag 2073, 88999 Kota Kinabalu, Sabah, Malaysia Tel: +6-088-320000, Ext: 3543	297
D4	Effect of Substrate Bias Voltage on The Surface Characterization of TiN coated HSS using CAPVD Technique	A. Mubarak, E. Hamzah, M. R. M. Toff	mubarak60@hotmail.com	Department of Materials Engineering, Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, MALAYSIA. Tel. +60-7-5534653, Fax. +60-7-5576820,	301
D5	Strain-Induced recrystallization kinetics of Nb-high strength low alloy steel during hot rolling	E.S.Siradj	siradj@metal.ui.ac.id	Department Metallurgy and Materials Engineering Faculty University of Indonesia. Kampus UI Depok 16424, Tel(021)7863510 Fax: 021 7872350	307
D6	The Properties and Kinetics of Carburized Duplex Stainless Steel	Nik Rozlin Nik Masdek, Iswadi Jauhari	nikrozlin@yahoo.com	Department of Mechanical Engineering, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia Tel: +019-3452554, Fax: +603-7967 5317	313
D7	The Pretreatments and Microwave Power Effects on Nucleation and Growth of Polycrystalline Diamond Coated Si ₃ N ₄	E. Hamzah A Purniawan., M. R. M. Toff	agung_purniawan@yahoo.com	Department of Materials Engineering, Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Malaysia Telp: +60-7-5534563, Fax: +60-7-5576820	319
D8	An experimental investigation of coated ceramic cutting tools when hard turning cold work tool steel	M. A. Kamely, M. Y. Noordin, V. C. Venkatesh, D. Kurniawan	kamelysita@yahoo.com	Faculty of Mechanical Engineering Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-7-5534697, Fax: +60-7-5566159,	325
D9	New Correlation between Karman Constant () and y-Intercept (B) for the Use of Clauser-Chart Technique to Estimate Wall Shear Stress	Sutardi	sutardi@me.its.ac.id	Mechanical Engineering Department, Faculty of Industrial Technology ITS, Surabaya, Indonesia, 60111	331

D Advanced Material and Materials Processing Technology

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
				Tel./Fac.: +62-31-5922941,	
D10	Cutting Force Predictions Models in End Milling Titanium Alloy Ti-6Al-4V	A.S. Mohruni, S. Sharif, M.Y. Noordin	mohrunias@yahoo.com	Department of Mechanical Engineering, Faculty of Engineering Sriwijaya University, 30662 Indralaya, South Sumatra Indonesia Tel: +60-7-5534850, Fax: +60-7-5566159	337
D12	Additives in EDM Dielectric	Norliana Mohd Abbas, Darius G. Solomon and Md. Fuad Bahari	chelorot@lycos.com	Faculty of Mechanical Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, Selangor Darul Ehsan, Malaysia Tel: +603-5543 5159, Fax: +603-5543 5160	343
D13	The effect of moisture on the extraction rate of palm pressed fiber residue using press machine	Zainoor Hailmee Solihin, Ahmed Jaffar	zainoor8@yahoo.com	Faculty of Mechanical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor Darul Ehsan, MALAYSIA Tel: 6012-2646482	349
D14	A Modeling and Simulation of Process Improvement in Polyurethane Injection Manufacturing Line	Noriah Yusoff, Ahmed Jaffar	norieyusof@yahoo.com	Faculty of Mechanical Engineering University Technology MARA, 40450 Shah Alam, Selangor, Malaysia	355
D15	Taguchi Methodology-based Approach to Precision grinding of Silicon.	Alao Abdur-Rasheed, Konneh Mohamed ²	abdur_rash@yahoo.co.uk	Manufacturing and Materials Engineering Department, International Islamic University Malaysia, 53100 Kuala Lumpur, Malaysia Tel: +6-0126865425, Fax: +60-3-61964568,	361
D16	Evaluation of a Material Flow Stress Model Adopted in Finite Element Modeling of Metal Cutting	M. N. Tamin, S. Izman, V. C. Venkatesh, T.T. Mon	pm033011@siswa.fkm.utm.my	Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia	369
D17	Microstructure and Creep Behaviour of AS-Cast Binary two phase Gamma TiAl.	E. Hamzah, ¹ M. Kanniah and ² M. Harun	esah@fkm.utm.my	Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor Malaysia Tel: 07-5534563 Fax: 07-5566159	375
D18	Performance of wiper coated carbide tool when turning hardened stainless steel	M. Y. Noordin, D. Kurniawan, S. Sharif, Y. C. Tang	noordin@fkm.utm.my	Faculty of Mechanical Engineering Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-7-5534697, Fax: +60-7-5566159	381

E Advanced Manufacturing

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
E1	Development Of Activity-Based Costing (ABC) Model In Demanufacture Cost Assessments	S.Mugeneswaran a/l Shanmugam , Awaluddin Mohd Shahraroun	muge@oum.edu.my	Faculty of Engineering & Technical, Open University Malaysia, Jalan Tun Ismail, 50480 Kuala Lumpur	387
E5	Condensation of Refrigerant R-22 in the Annulus of Horizontal Double-Tube Condenser with an Enhanced Inner Tube	R. Tiruselvam, Mohd. Zainal Yusof, Vijay R. Raghavan	vijay@kuittho.edu.my	Fakulti Kejuruteraan Mekanikal & Pembuatan Kojel Universiti Teknologi Tun Hussein Onn, Parit Raja Tel:+60-125698389	393
E6	Studies on Modelling of a Swirling Fluidized Bed	Mohammad Kamil, Mokhtar Yusoff, Vijay Raghavan	vijay@kuittho.edu.my	Department of Plant and Automotive Faculty of Mechanical and Manufacturing Engineering Kolej Universiti Teknologi Tun Hussein Onn, 86400 Parit Raja, Johor, Malaysia Tel: +60-7-4537794, Fax: +60-7-4536080,	399
E7	Analysis of Flow and Heat Transfer over Louvered Fins in Compact Geometries	Yong Kok Wee, Vijay R. Raghavan	vijay@kuittho.edu.my	Kolej University Teknologi Tun Hussein Onn, 86400 Parit Raja, Batu Pahat, Johor, Malaysia.Tel:+60-126793735,	405
E8	Experimental Study on Manoeuvring Characteristics of a Planing Hull	A. Haris Muhammad, Adi Maimun, Omar Yaacob, Agoes Priyanto	andi_haris@ft.unhas.ac.id	Faculty of Mechanical Engineering Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-7-5535706,	411
E9	Feasibility Study of Ultrasonic Transducers as Altitude and Attitude Sensors for Flight Control System	T. Indriyanto and U. M. Zaeny	t.indriyanto@ae.itb.ac.id	Aerospace Engineering, Faculty of Industrial Technology Institute of Technology Bandung, Bandung, Indonesia Tel: +62-22-2504529, Fax: +62-22-2534164,	471
E10	Development of Cavity Defect in Forward Extrusion	Tri Widodo Besar Riyadi, John Atkinson, Waluyo Adi Siswanto	tri_wbr@yahoo.com	Department of Mechanical Engineering Muhammadiyah University of Surakarta, Pabelan, Surakarta, Indonesia. Tel: +62-271-717417 ext 222, Fax. +62-271-715448,	423
E11	Numerical Simulation of Jet Impingement Cooling on a Smooth Concave Surface	Suzairin bin Md Seri and Vijay R. Raghavan	vijay@kuittho.edu.my	Faculty of Mechanical and Manufacturing Engineering Kolej Universiti Teknologi Tun Hussein Onn(KUiTTHO)86400 Parit Raja, Batu Pahat, Johor, Malaysia Tel: +60-7-4537794, Fax: +60-7-4536080	429
E13	Particle Swarm Optimization Neural Network Based Modelling of Vehicle Suspension System	Gigih Priyandoko Musa Mailah Hishamuddin Jamaluddin	gigihp@gmail.com	Department of Applied Mechanics, Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor Bahru, Malaysia	435
E14	Preliminary study of active noise control method for jet engine noise using secondary multipole resource	Muhammad Kusni, Rudyansyah	kusni@ae.itb.ac.id	Study Program of Aerospace Engineering, Faculty of Industrial Technology Bandung Institute of Technology, Bandung, Indonesia Tel: +62-22-2504529,	441

E Advanced Manufacturing

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
E15	Optimum Number of Stages of the New Multi-Stage Symmetrical Wobble Plate Compressor	Ardiyansyah Syahrom, Md. Nor Musa , Wan Ali Wan Mat, Ainullofi Abdul Latif	ardiyansyah@utm.my	<i>Faculty of Mechanical Engineering Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-7-5534878, Fax: +607-5566159,</i>	447
E21	Ratio control of an electromechanical dual acting pulley – continuously variable transmission (EMDAP-CVT) system using PD fuzzy logic controller	Bambang Supriyo, Kamarul Baharin Tawi, Hishamuddin Jamaluddin and Sugeng Ariyono	bsupriyo7763@yahoo.com	<i>Department of Electronic Engineering, Politeknik Negeri Semarang Jl. Prof.H. Sudarto, S.H. Tembalang, Semarang 50329, Indonesia</i>	455
E22	Micro/Meso Mechanical Manufacturing (M ⁴): Opportunities and Challenges - A Review	R. Mehruz, M. Y. Ali*	mmyali@iiu.edu.my	<i>Department of Manufacturing and Materials Engineering Faculty of Engineering, International Islamic University Malaysia 53100, Jalan Gombak, Selangor Darul Ehsan, Malaysia</i>	463
E23	Development of Symmetrical Double Sided Linear Wobble Plate Compressor for Bus A/C System	Mohd Suffian, Imran Sairaji, Md. Nor Musa	mdnor@fkm.utm.my	<i>Faculty of Mechanical Engineering UTM Skudai, Johor. Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-7-5534567, Fax: +60-7-5536688,</i>	469

F Petroleum and Natural Resources Process

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
F2	Degradation of phenolic compounds by indigenous isolated microbes	M. Mailin, R. Firdausi	jh_mailin@yahoo.com	Department of Bioprocess Engineering, Faculty of Chemical & Natural Resources Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor.	473
F3	Optimization of Process Conditions for Citric Acid Production Utilizing STP Sludge as Substrate by Liquid State Bioconversion	<u>Md. Zahangir Alam</u> , Parveen Jamal and Rosnani Saari	zahangir@iiu.edu.my	Bioenvironmental Engineering Research Unit (BERU), Department of Biotechnology Engineering, Faculty of Engineering, International Islamic University Malaysia (IIUM), Jalan Gombak, 53100 Kuala Lumpur, Malaysia. Tel: +603-6196457, Fax: +603-61964442	479
F4	Feed mill from Solid Waste of Flushing Resulted from Leather Processing Industry	<u>Elly Agustiani</u> , Soeprijanto, Danawati H.P	cmmap@sby.centrin.net.id	Department of Chemical Engineering Sepuluh Nopember Institute Technology of Surabaya, Sukolilo, Surabaya, Indonesia Tel: +62-81-23027006, Fax: +62-031-5935368	485
F8	The Effect of Strach Pretreatment on Cyclodextrins Production	<u>A.M. Mimi Sakinah</u> , Rosli Md Illias, A.F. Ismail, Osman Hassan, A.W Zularisam	mimi@kuktem.edu.my	Faculty of Chemical Engineering and Natural Resources, University of College Engineering and Technology of Malaysia (KUKTEM), Kuantan Pahang, Malaysia Tel: +60-095492322, Fax: +60-095492322,	489
F9	Autothermal Reforming of Methane using NiO-CoO/MgO Catalyst	Tutuk D. Kusworo · A.R. Songip and N.A. Saidina Amin	jokotutuk@yahoo.com	Department of Chemical Engineering Universiti of Diponegoro, Semarang, Indonesia Jl. Prof. Sudharto Tembalang, Semarang Indonesia	493
F11	Statistical Screening of Medium Components for Extracellular Expression of Recombinant Cyclodextrin Glucanotransferase in <i>Escherichia coli</i>	<u>Lo Po Kim</u> , Tan Cheau Yih, Arshad Ahmad, Rosli Md. Illias	lopokim@time.net.my	Department of Bioprocess Engineering Faculty of Chemical and Natural Resources Engineering Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia Tel: 60-7-5535472, Fax: +60-7-5581463,	499
F14	Application of Sol-Gel Coating for Solid Phase Microextraction Fiber to the Analysis of Organophosphorus Pesticides	Wan Aini Wan Ibrahim, <u>Nor Fairolzukry Ahmad Rasdy</u> , Mohd Marsin Sanagi, Ahmedy Abu Naim	wanaini@kimia.fs.utm.my	Department of Chemistry, Faculty of Science Universiti Teknologi Malaysia, 81310 UTM, Johor, Malaysia Tel: +60-7-5535914, Fax: +60-7-5536688	505
F16	Electron Microscopes Analysis of Home Grown Multi-Walled Carbon Nanotubes (MWNTs)	Tee Jia Chee, Nor Aziah Buang, Ahmad Fauzi Ismail	noraziah@kimia.fs.utm.my	Department of Chemistry, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia.	509

F Petroleum and Natural Resources Process

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
				Tel: +60-7-5534123, Fax: +60-7-5534518	
F17	Residential Wastewater as a Source of Competent Phenol Degrading Consortia	<u>Norhani Jusoh</u> and Firdausi Razali	firdaus@fkkksa.utm.my	Department of Bioprocess Engineering, Faculty of Chemical & Natural Resources Engineering Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-7-5535805, Fax: +60-7-5581463	515
F18	Vapor Liquid Equilibrium of Ethanol Water System Effect of Ultrasonic Waves	Siti Kholijah Abdul Mudalip, Adnan Ripin, Rosli Mohd Yunus	adnan@fkkksa.utm.my	Chemical Engineering Department, Faculty of Chemical & Natural Resources Engineering, Universiti Teknologi Malaysia, 81300 Skudai, Johor Bahru, Malaysia. Tel: +60-7-5535690, Fax: +60-7-5581463	521
F19	The Use of Surfactant Modified Zeolite Y as a Sorbent for Arsenic Species in Aqueous Solution	<u>Nik Ahmad Nizam</u> and Alias Mohd Yusof	Alias@kimia.fs.utm.my	Department of Chemistry, Faculty of Science Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-7-5534500, Fax: +60-7-5566162,	527
F20	Immobilization of Glucose Oxidase and Ferrocene Redox Polymer in cross-linked poly (vinyl alcohol) with bovine serum albumin as Protein Stabilizer	<u>Norhana Jusoh</u> , A Abdul Aziz	azila@fkkksa.utm.my	Faculty of Chemical and Natural Resources Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia. Tel: +60-7-5535621, Fax: +60-7-5581463	533
F21	Stability and Demulsification of Water-in-Crude Oil (w/o) Emulsions Via Microwave Heating	Abdurahman.H.Nour and Rosli, Mohd. Yunus	nour2000_99@yahoo.com	Faculty of Chemical and Natural Resources Eng. Universiti Teknologi Malaysia 81310 Skudai, Johor-Malaysia.	539
F22	Supercritical Fluid Carbon Dioxide for the Extraction of Polychlorinated Biphenyls in Sewage Sludge	Umi K. Ahmad and Ahmad Zamani Ab Halim	umi@kimia.fs.utm.my	Department of Chemistry, Faculty of Science, Universiti Teknologi Malaysia 81310 UTM Skudai, Johor, Malaysia. Tel: 07-5534522	545
F23	Optimization of Headspace SPME using Lab-made C8-coated Fiber for the Forensic Analysis of Arson Accelerants in Fire Debris	Umi K. Ahmad, Abdul Rahim Yacob, <u>Geetha Selvaraju</u>	umi@kimia.fs.utm.my	Department of Chemistry, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +07-5534522, Fax: +07-5566162	551
F25	Principles of novel continuous ion exchange plate column for Heavy metal removal	Idral and Saidi, H.	idra_amri@yahoo.com	Membrane Research Unit, Department of Chemical Engineering, Universiti Teknologi Malaysia, Jln. Semarak 54100, Kuala Lumpur, Malaysia Tel: +60-3-26154823,	557

F Petroleum and Natural Resources Process

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
				Fax : +60-3-26914427	
F26	Preparation and Characterization of Delaminated Zeolite, ITQ-6	Noor Aishikin M. Y. and Zainab R.	zainab@kimia.fs.utm.my	Department of Chemistry, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia. Tel: +60-75534491, Fax: +60-7-5566162,	563
F27	Effect of Temperature on Catalytic Cracking of Palm Oil to Gasoline Over Copper-HZSM-5	Tirena Bahnur Siregar	tirenasiregar@yahoo.com	Faculty of Chemical and Natural Resources Engineering Universiti Teknologi Malaysia, 81310 UTM, Skudai, Johor, Malaysia	569
F30	Characterization of Acid-treated Carbon nanotubes (CNTs) Synthesized using MgO and CaCO ₃ Supports	Mohamed Nuruddin bin Mohamed Nasir, Nor Aziah bte Buang	noraziah@kimia.fs.utm.my	Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia. Tel. No. (Office): 07- 5533333 (ext: 34123) Fax: 07-5566162	575
F32	Synthesis of highly reactive electrogenerated Zinc for oxidative addition of organic halides	Murni, S, Aishah, A. J, Sugeng, T., Muhd. Nazlan, M. M. and Masao, T.	aishah@fkkksa.utm.my	Faculty of Chemical and Natural Resources Engineering, ² Ibnu Sina Institute for Fundamental Science Studies, ³ Department of Chemistry, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Bahru, Johor. Tel: +60-7-5535523, Fax: +60-7-5581463,	579
F33	Oil Water Contact Analysis and Hydrocarbon Saturation Estimation Based on Well Logging Data	Jarot Setyowiyoto and Ariffin Samsuri	j_setyowiyoto@yahoo.com	Universitas Gadjah Mada Indonesia, Universiti Teknologi Malaysia Phone: +60167 285 312	583
F34	Factors controlling sediment budget at river mouths in the south coast of central Java Indonesia	Hendra Amijaya, Sugeng S. Surjono	hamijaya@ugm.ac.id	Department of Geological Engineering, Gadjah Mada University Jalan Grafika 2, Yogyakarta, Indonesia Tel: +62-274-6491380/- 513668, Fax: +62-274-513668	589
F35	Decolourization of Azo Dye Direct Blue 15 Using Batch Culture of <i>Klebsiella</i> sp	Chun Shiong Chong, Zaharah Ibrahim, Madihah Md Salleh, Noor Aini Abdul Rashid, Adibah Yahya, Wui Jin Wong	tcchiong80@yahoo.com	Department of Biology, Faculty of Science Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia Tel: +60-7-5532528, Fax: +60-7-5566162	595
F41	Microbial Fuel Cell Operating on Cellulose for Electricity Generation	Aslizah, M.A, Adibah, Y., Zaharah, I.	aslizah_n@yahoo.com	Biology Department, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +607-5532528, Fax: +607-5566162	601
F42	Contribution from the Domestic Sector Considering various Scenarios and by using effective tool (SAP Energy Rating) to meet	S R Samo BE MEng PhD MPEC MIEP	srsamo@yahoo.com	Department of Energy and Environment Engineering, QUEST, Nawabshah Sindh, Pakistan	607

F Petroleum and Natural Resources Process

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
	the Kyoto Commitment			Tel 0092 244 9370361, cell 0333 7010738 fax 0092 244 9370367	
F44	Influence of Chlorobenzene Levels on Bioremediation by Microbial Consortia from Residential Wastewater	<u>Noor Hafiza Harun</u> , Firdausi Razali	fiexhar@yahoo.com	Department of Bioprocess Engineering Faculty of Chemical and Natural Resources Engineering University of Technology Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-7-5535513, Fax: +60-7-5581463	613
F45	CO-Generation of synthesis gas and higher hydrocarbons over catalytic dielectric barrier discharge plasma reactor	Istadi, and Nor Aishah Saidina Amin	istadi@tekim.ft.undip.ac.id	Chemical Reaction Engineering and Catalysis (CREC) Group, Department of Chemical Engineering, Diponegoro University, Jln. Prof. Sudharto, Semarang, Indonesia 50239	619
F46	Production and Characterization of crude chitinase from <i>Trichoderma virens</i>	Azaliza Safarida Wasli, Madiah Md Salleh, Rosli Md Illias	azasafarida@yahoo.com	Biology Department, Faculty of Science University Technology Malaysia, 81310, Skudai, Johor Darul Takzim	625
F47	Degradation of Thiosulfate by Sulfide-Oxidizing Enzyme Produced by Bacteria Locally Isolated from Effective Microorganism Active Solution (EMAS)	<u>Nur Suhanawati Ashaari</u> , Madiah Md. Salleh and Rosli Md. Illias	suhana_bio@yahoo.com	Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia. Tel: +60-7-5532526	631
F48	Research Challenges on Natural Gas Treating Process	Iwan Ratman and A.F. Ismail	Iwan.ratman@gmail.com	Brunei Shell LNG Sdn Bhd	637
F49	Kinetic Modeling of Methane Conversion in the Presence of Co-Feeding to Gasoline Hydrocarbon Products over W/HZSM-5 Catalyst	<u>Kusmiyati</u> and Nor Aishah Saidina Amin	kmy_ums@yahoo.com	Department of Chemical Engineering Muhammadiyah University of Surakarta, Pabelan, Surakarta, Indonesia Tel: +62-271-717417, Fax: +62-271-7174174,	645
F50	Isolation, Screening and Characterization of Soluble Exopolymer-Producing Bacteria For Enhanced Oil Recovery	<u>Munirah Tharek</u> , *Zaharah Ibrahim, S.Hasila Hamzah, Noraha Markum, Aslizah Mohd Aris, Fareh Nunizawati Daud, Madiah Md Salleh, Adibah Yahya, Liew Chong Wai, Nozieana Khairuddin, Rosli Illias, Mohamad Ismail Omar, Khor Siak Foo, Ezrin Johanna Elias, Scott Bailey	munirahtharek@yahoo.com	Department of Biology, Faculty of Science Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-7-5532526, Fax: +60-7-5536688	649

G Architecture

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
G1	The effectiveness of environmental impact assessment (EIA): A case study on housing projects.	ROSLINDA BINTI ALI AND MANSOR BIN IBRAHIM	roslanda1110@yahoo.com	<i>Kulliyah of Architecture and Environmental Design International Islamic University Malaysia, P.O. Box 10, 50728 Kuala Lumpur.</i>	655
G2	Thermal Ventilation Performance of Malaysia's Single Storey Terrace House	Agung Murti Nugroho ¹ , Mohd Hamdan Ahmad ²	sasimurti@yahoo.co.id	<i>Faculty of Built Environment Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-127374703,</i>	681
G3	Mosque Architecture in Malaysia: Classification of Styles and Possible Influence	Mohamad Tajuddin Mohamad Rasdi, Nangkula Utaberta	Nangkula_arch@yahoo.com	<i>Dept. of Architecture, Faculty of Built Environment Universiti Teknologi Malaysia UTM Skudai, Johor Bahru, 803</i>	687
G4	Redevelopment The Malay Agriculture Settlement (MAS) Areas of Kampong Baru Using Land Readjustment Technique	Prijono Nugroho Djojomartono, Djurdjani, Ismail Omar, Asiah bte Othman		<i>Department of Land Administration and Development Faculty of Geoinformation Science and Engineering Universiti Teknologi Malaysia</i>	693
G5	TRANSFORMATION: Impact of the Federal Capital on Gbagyi (Gwari) Built Environment, 1976-2005	Moukhtar M. Mai, & Dr. Suhanna Shamsuddin	mouktarmai@yahoo.com	<i>Dept. of Architecture, Dept. of Faculty of Built Environment Universiti Teknologi Malaysia Tel. +60-146262418</i>	699
G6	DEVELOPMENT OF VEGETATION AND LAND USE BY REMOTE SENSING METHOD IN WADI TUBAN ABYAN AREA - YEMEN	A. A. Almhab, I. Busu , N. Ibrahim	ayoub.almhab@fkg.utm.my	<i>Department of remote sensing Faculty of Geoinformation Science & Eng University Teknologi Malaysia</i>	705

H Economy

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
H2	Impacts of Electronic Data Interchange (EDI) Benefits on Firm Performance: Evidence from Malaysian Manufacturers	Gengeswari, K., and Abu Bakar, A.H.	k.gengeswari@gmail.com	<i>Department of Management, Faculty of Management and Human Resource Development Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia Tel: +6075531899, Fax: +6075531908,</i>	711
H3	Openness and Productivity Growth in Malaysia	Siti Sumiati ¹ and Assoc Prof. DR Azmi Abd. Rahman ²	siti_sumiati@yahoo.com	<i>Faculty of Management and Human Resources Development Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Telp: +60-7-5537447,</i>	717

Poster Presentation.

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
A7*	Natural Organic Matter (NOM) Fouling in Surface Water Treatment	A.W. Zularisam, A.F. Ismail, M.R. Salim	afauzi@utm.my	Environmental Engineering Focus Group, FKA, Kolej Universiti Kejuruteraan & Teknologi Malaysia, KUKTEM, Gambang, Pahang, Malaysia	725
A8*	The Characteristics of the Bus Route Length and Fleet in Jakarta	Leksmono Suryo Putranto	lexy@tarumanagara.ac.id	Department of Civil Engineering Tarumanagara University, Jl. Let. Jen. S. Parman No. 1, Jakarta 11440, Indonesia Tel: +62-21-5672548, Fax: +62-21-5663277,	733
A23*	A Quantitative Approach: Information Technology Infrastructure Capability Framework in Malaysian Construction Consulting Companies	M. Z. Abd. Majid, Amran Rasli and Ade Asmi	a_asmi99@yahoo.com	Department of Civil Engineering, Universitas Indonesia	737
B16*	A Feature-based Music Score Recognition Method	Agus Harjoko ¹ and I Gede Ngurah Agung Darmawan	aharjoko@ugm.ac.id	Electronic and Instrumentation Lab., FMIPA, Gadjah Mada University, Yogyakarta – Indonesia	745
F7*	Oxygen Enrichment Properties of Polyethersulfone/Polyimide Blends-Zeolite 5A Mixed Matrix Membranes	Tutuk D. Kusworo A.F. Ismail and A. Mustafa	afauzi@utm.my	Membrane Research Unit Faculty of Chemical and Natural Resources Engineering Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia	751
F10*	The effect of cross flow velocity on the permeate flux during cyclodextrins bioseparation	A.M. Mimi Sakinah, Rosli Md Illias, A.F. Ismail, Osman Hassan, A.W Zularisam	mimi@kuktem.edu.my	Faculty of Chemical Engineering and Natural Resources, University of College Engineering and Technology of Malaysia (KUKTEM), Kuantan Pahang, Malaysia Tel: +60-095492322, Fax: +60-095492322,	757
F24*	Modeling, simulation, and optimization of Continuous Ion Exchange (CIXC) for heavy metal removal from industrial wastewaters	Idral and Saidi, H.	idra_amri@yahoo.com	Membrane Research Unit, Department of Chemical Engineering, Universiti Teknologi Malaysia, Jln. Semarak 54100, Kuala Lumpur, Malaysia Tel : +60-3-26154823, Fax : +60-3-26914427	761
F51*	Preparation of Polysulfone Ultrafiltration membranes via Microwave Technique	Ani Idris, Iqbal Ahmad	ani@fkkksa.utm.my	Department of Bioprocess Engineering Faculty of Chemical and Natural Resources Engineering Universiti Teknologi Malaysia, 81310 Skudai, Johor Tel: +607-5535603, Fax: 607-5581463	767
F52*	Influence of bore fluid flow rate to ultrafiltration hollow fiber membrane performance	Ani Idris, Sze-Yean Lee	ani@fkkksa.utm.my	Department of Bioprocess Engineering, Faculty of Chemical and Natural Resources Engineering Universiti Teknologi Malaysia 81300 UTM Skudai, Johor Bahru. Tel: +607-553560, Fax: +607-5581463	771

Additional Papers and Poster

A Construction and Build Environment

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
A19	Value Management in Construction Industry	Bachan Singh, Abdul Rahim Abdul Hamid, Ng Kim Lai	bachan19@yahoo.com	Faculty of Civil Engineering Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-7-5531538, Fax: +60-7-5566157,	803

B Information and Communication Technology

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
B19	Digital Watermarking, Challenges and Applications	Akram M. Zeki and Azizah A. Manaf	akramzeki@yahoo.com	Faculty of Computer Science and Information System Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia Tel: +60-19-6340683,	809
B22	Measurement of Health Risk Events using RISKAS v2 for Arsenic Contamination	Sri Adelila Sari , Zaini Ujang, Umi Kalthom Ahmad, Lukman Hakim Ismail	adelila@yahoo.com	Department of Chemistry, Faculty of Science ² Institute of Environmental and Water Resource Management Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia. Tel: +607-5531578, Fax: +607-5531575,	813

C Electrical and Electronics

No.	Title	Authors	Corresponding e-mail	Corresponding Address	Page No
C8	NN with DTW-FF Coefficients and Pitch Feature for Speaker Recognition	Rubita Sudirman, Sh-Hussain Salleh, Shaharuddin Salleh	rubita@fke.utm.my	Faculty of Electrical Engineering Faculty of Science Universiti Teknologi Malaysia, 81310 Skudai, Johor Tel.: 607-553 5738, Fax:607-553 5681	775

E Advanced Manufacturing

E12	Vibration Characteristics of Shape Memory Alloy Hybrid Composite Plate	W.C. Tan, A. Murni, M. I. Ghazali and S. Jamian	tweechoon@hotmail.com	Faculty of Mechanical and Manufacturing Engineering Kolej Universiti Teknologi Tun Hussien Onn Beg Berkunci 101, 86400 Parit Raja, Batu Pahat, Johor Darul Takzim, MALAYSIA	817
E19	Proposed Mechanism of Indigenously Electromechanical Dual Acting Pulleys CVT (EMDAP CVT) for Automotive Application	S. Ariyono, K B Tawi, B Supriyo, N Abu Husain	abusafani@yahoo.co.uk	Drivetrain Research Group Automotive Development Center Faculty of Mechanical Engineering Universiti Teknologi Malaysia (UTM)	781

Additional Papers and Poster

F Petroleum and Natural Resources Process

F12	Thermal Behaviour of Coal /Biomass Blends During Co-Pyrolysis	Ahmad Hussain, Farid Nasir Ani, Amer Nordin Darus	ahmadutm@gmail.com	<i>Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310, Skudai, Johor Tel: +607-5534785, Fax: +607-5566159,</i>	789
F43	Pineapple Waste Juice Utilization for Ethanol Production by <i>Saccharomyces Cerevisiae</i>	<i>Abdullah, Yefta, Popi</i>	abd_busair@yahoo.com	Department of Chemical Engineering, Faculty of Engineering, Diponegoro University, Semarang. Tel: +62-24-7460058	829
F53	Optimization of Cellulase and Lignin peroxidase Production by Locally Isolated <i>Aspergillus</i> spp. Using Oil Palm Empty Fruit Bunch (OPEFB)	Salwa Mohamad, Zaharah Ibrahim, Madiah Md Salleh	cellewa@yahoo.com	<i>Biology Department, Faculty of Science, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia Tel: +60-7-5532528</i>	823

Poster Presentation.

F13	Investigation on Co-Combustion Characteristics of Palm Shell Waste and Coal in Circulating Fluidized Bed Combustor	Ahmad Hussain, Farid Nasir Ani, Amer Nordin Darus	ahmadutm@gmail.com	<i>Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310, Skudai, Johor Tel: +607-5534785, Fax: +607-5566159</i>	795
-----	--	--	--	--	-----

Papers



Need of Sustainability in Construction Industry of Pakistan

S.Mubin ¹

¹ Faculty of Civil Engineering

University of Engineering and Technology, Lahore, Pakistan

Tel and Fax: +92-42-6829202 web: www.uet.edu.pk, E-mail: sajjadmubin@uet.edu.pk

Abstract

Sustainability' is becoming a central concern for all of us that rising populations and economic development are threatening a progressive degradation of the earth's resources. Over the next three decades the world's population is expected to grow by some 2.2 billion people means ever-increasing demand for new structures and infrastructure. As a whole this growth of population is very sharp in Asian and particularly in south Asian countries. The Construction, maintenance and use of buildings drain energy and natural resources and impact substantially on our environment, atmosphere and ecosystem. Construction industry is one of the largest in terms of economic expenditure, volume of raw materials and natural resources consumption, employment generation and environmental impacts, etc. Large variety of materials are manufactured and consumed in the construction industry. Bulk energy and natural resources are consumed in production of construction materials like bricks, cement, Steel, aluminium and glass. How to make the construction more sustainable, environment friendly, cost effective, durable and with least operation and maintenance costs? This important issue has been focused and effort has been made to present economical and cost effective solutions/models with reference to building construction in south Asian countries generally and particularly in Pakistan.

Keywords:

Sustainability, Construction, Low Cost Housing, Green Building Construction, Pakistan.

1. Introduction

The paper addresses certain issues pertaining to the energy, environment, alternative building technologies and sustainable building construction. Buildings account for one-sixth of the world's fresh water withdrawals, one-quarter of its wood harvest, and two-fifths of its material and energy flows [1] Brief history of developments in building materials is discussed. Energy Consumption in manufacture and transportation of some common and alternative building materials and the implications on environment are presented. Brief details of some of the energy-efficient alternative building technologies developed by CWHR and other research institution are provided. Impacts of alternative building technologies on energy and environment are discussed. Some thoughts about utilizing industrial and mine wastes as well as recycling of building wastes for meeting the demand for buildings in a sustainable fashion have also been presented. Building materials and technologies, and building practices have evolved through ages. Quest for durable building materials is an ongoing phenomenon ever since man started construction activity. Brick burning represents one of the earliest examples of using energy (other than animate energy) to manufacture durable building materials from the soil/earth. Firewood was the main source of energy for burning bricks. Use of metal products represents the next energy consuming, manufactured material for the construction, after bricks. Then comes the manufacture of

lime and lime-based products. Burnt bricks, metal products and lime are the manufactured materials (using external energy sources) commonly used for construction by our ancestors for a very long period of time. Discovery of natural inorganic binders like pozzolanic materials lead to the use of lime-pozzolana (LP) cement for construction purposes. Experience of using LP cement paved the way for the invention of Portland cement in 1824. Portland cement and steel (concrete) brought revolutionary changes in the construction practices from early part of 20th century. Then plastics and plastic products entered the construction industry. Chronological sequence of developments in building materials and products is listed in Table 1.

Material	Period
Mud, stones, wood	Prior 8000 BC
Sun aried bricks	6000 BC
Pottery products	4000-8000 BC
Burnt bricks	4000 BC
Lime	3000 BC
Glass	1300 BC
Iron products	1350 BC
Lime-pozzolana cement	300 BC-476 AD
Aluminum	1808 AD
Portland Cement	1824 AD
Plastics	1862 AD

Table 1. Historical development of Building Material

Bricks, cement, steel, aluminium, plastic products, paints, polished stone, ceramic products, etc. are the commonly

used materials of construction today. These materials are energy intensive and are transported over large distances before being used for construction. The following points require attention, regarding the use of modern building materials: Energy consumed in the manufacturing processes – energy intensity; Problems of long distance transportation; Natural resources and raw materials consumed; Recycling and safe disposal; Impact on environment, and Long-term sustainability. Thus the issues related to energy expenditure, recycling, biodegradable, environmental and sustainability with respect to future demand need to be addressed during the manufacture and use of any new building material. Some issues on energy, environment, sustainable construction and buildings

Material	Volume of material manufactured per annum (2003)	Thermal energy (MJ per Kg)	Total Energy GJ
Brick	160×10^9 Nos	1.6	710×10^6
Cement	120×10^6 tonnes	5.1	500×10^6
Aluminum	1×10^6 tonnes	250.6	200×10^6
Structural steel	13×10^6 tonnes	47.0	510×10^6

Table 2. Energy consumption by construction material[2]

Environmental Impact of Buildings Percentage of U.S., Annual Impact

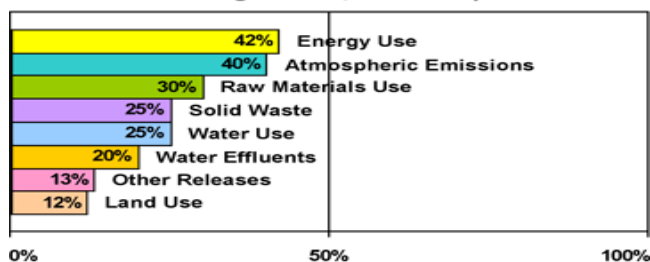


Figure 1. Environmental impact of Building [3]

2. Sustainable Construction in Pakistan

Pakistani construction industry is one of the largest in terms of economic expenditure, volume of raw materials/natural resources consumed, volume of materials and products manufactured, employment generated, environmental impacts, etc. Large variety of materials are manufactured and consumed in the construction industry. Production levels and energy expenditure of some of the building materials consumed in bulk quantities are given in Pakistan is at the initial stage of realization of sustainable construction/development. There is a long way to go for getting sustainable awareness which is again dependent of social system, literacy rate and governmental policies and involvement. There are a lot of organizations working towards attaining sustainable development. PCRET and CWHR are working under Ministry of Science and Technology. PCRET (Pakistan Council for Renewable Energy and Technology) has finished many projects in rural

areas development with construction of bio gas plants all Pakistan basis. The end users are the poor people living at poverty or below poverty level. These bio gas plants are operated with animal waste and being used successfully replacing wood by gas and reducing the environmental impact of operations of buildings. PCRET is these days actively busy in developing other modes of renewable energy from sun and wind. Many solar plants are successfully working in Karachi. However, studies are being conducting to capture the wind energy of the costal city of Karachi for production of electricity. CHWR(Council for works and housing research) is engaged in developing low cost construction material, techniques and processes. It has developed many products which are patient and successfully being used in the market. That includes cement stabilized block, low cost roofing system, ferrocemnt manholes etc. It also provided 100 ferrocemnt benches installed in famous Bagh-e-Jinnah, Karachi. It was first experience of its kind. besides, it developed different types of machines for improvement of construction processes in reduction of energy during fabrication of consruction material. Besides that, Agha khan Planning and Building Services (AKPBS) is working in northern areas of Pakistan on rural development projects, providing residential units and safe drinking water to the population. In December 2005, The Aga Khan Planning and Building Service in Pakistan (AKPBS,P) was named the 2005 winner of the US \$1 million Alcan Prize for Sustainability by the Prince of Wales International Business Leaders Forum (IBLF).[3]



Figure 1 A water filtration system above Altit, in the Northern Areas of Pakistan.

It received the award for its efforts to improve Pakistan's built environment and water and sanitation facilities. Launched in 1997, the programme's aim is to produce low cost residential units and to reduce the risk of water-borne diseases through the provision of potable water as well as improved hygiene and sanitation practices. The success of this programme has led to a partnership between AKPBS,P and the Pakistan Poverty Alleviation Fund. As a result, the programme is being replicated in several other areas with the support of donors and partners. In Pakistan the estimated time for attaining stage-3 (shown in Fig-1) is almost 80 years. However it can be reduced by improving factors affecting sustainability mentioned above. Fig 1 shows the characteristic curve between Sustainable Development v/s Time.

3. How the Sustainable Construction can move towards Sustainable Development?

The starting point for all members of the construction industry that wish to approach sustainability as a business opportunity - from big developers to small companies and those who refurbish existing buildings - must be to re-think their operations in four key areas:

Towards Sustainability

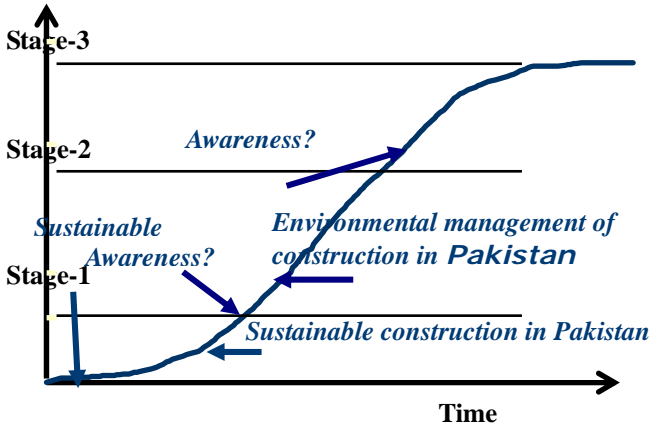


Figure 5 Trend of sustainable construction in Pakistan

- **Energy:** reducing energy consumption, being more energy efficient and using renewable energy and alternative technology'.
- **Materials:** Choosing, using, re-using and recycling materials during design, manufacture, construction and maintenance to reduce resource requirements.
- **Waste:** Producing less waste and recycling more.
- **Pollution:** Producing less toxicity, water, noise and spatial pollution.

4. Attainment of Sustainability in building construction

For achieving sustainability in building construction it is first necessary to identify the different phases of construction. For ideal results complete life cycle of the project should be studied with a view point of continuous improvement of each phase with reduction of resources and minimization of its impact on environment. Target should be not only reduce the over all cost but also the reduction of environmental impact due to construction and operation.

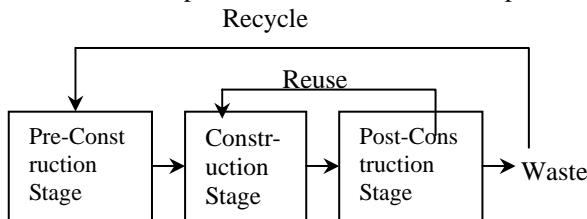


Figure 3 Stages of construction

Construction phases are further subdivided into two main categories

4.1. Sustainability with respect to construction process

Preconstruction phase with respect to construction consists of Concept clearance and feasibility, Planning and economical analysis, engineering design and Construction phase consists of following stages. Acquisition of Land,

Construction process and installation while post construction stage consists of operation and maintenance, dismantling and disposal of waste after the completion of design life. However the construction can be made more sustainable considering every phase of construction process

4.1.1. Project Feasibility and Planning

During feasibility and planning process cost may be reduced by proper planning. Project planning is very important in this regard. While considering Planning and feasibility phase, engineers and planners should take care of the environmental impact, reduction in energy use (in terms of material and transportation) , economic analysis, low cost , safety and durability. All above mentioned items should be considered in all phases of construction process i.e. feasibility, design, construction process, operation and maintenance, Demolition and reconstruction.

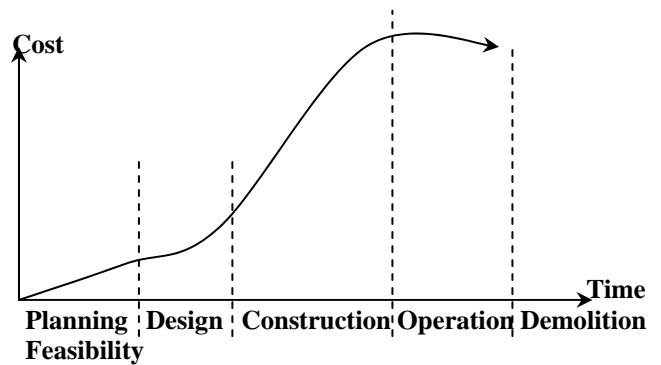


Figure 4. Life-Cycle of a construction project

Project is usually economically justified for large scale building projects with different analysis methods (IRR, ERR, AW, PW, B/C ratio method and etc), but while comparing alternatives usually the cost of environmental damages is ignored or over looked. Project Planning and construction management is very important in improvement of process and reducing the extra indirect costs involved. This is done by breaking structure in small activities (WBD) and scheduling them by identifying critical activity so that wastage of resources may be ignored.

If a housing scheme is going to be constructed and instead of rectangular plot, square plot gives the equivalent area that of rectangle but 20 percent lesser perimeter. In this case over all cost of construction may be reduced by the 30 percent shown in Fig-5 below.

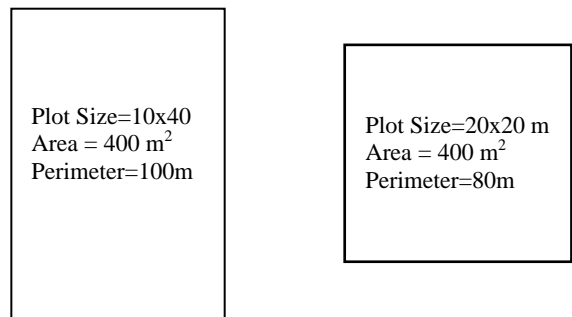


Figure 5 shapes of plot effect the cost of construction

In addition two storey houses gives the lesser cost compare to single unit. There is trend in Pakistan lowering the ceiling height upto 10.5-11 ft (3.5m) as compare to 13-13.5ft 30 years back. This saves the brick work.

4.1.2. Design

Design phase is divided on two main sub phases Structural Design and Architectural Design

In design phase cost of construction can be largely reduced in two ways. By lowering factors of loads and by innovation in design techniques. Concrete Reinforced with FRP Bars introduced by the American Concrete Institute (ACI). For new construction, FRP bars have been used as the internal reinforcement in concrete members to replace conventional steel rebars for a host of reasons. The principles for design and construction were first established and proposed to industry in 2001. in general, FRP materials are anisotropic and are characterized by high tensile strength with no yielding only in the direction of the reinforcing fibers. This anisotropic behavior affects the shear strength and dowel action of FRP bars, as well as their bond performance. Design procedures account for a lack of ductility in concrete reinforced with FRP bars [8]. An FRP-RC member is designed based on its required strength and then checked for serviceability and ultimate state criteria (e.g., crack width, deflection, fatigue and creep rupture endurance). In many instances, serviceability criteria may control the design. Similarly, fiber reinforcement bars can be used in flexure members under low loads. Ferro cement techniques may be introduces. But all that need a big change in concepts and cultures. Similarly Tru-Frame Steel Truss can be introduced for improvement of work pace and efficiency particularly after earthquakes, repair and replacement of structural elements is time consuming and expensive. This new truss system contains several diagonal elements designed to absorb seismic energy and yield at prescribed levels. After a quake, the diagonals are easily accessible for repair or replacement. Erection time saving can vary based on quantity and size of frames along with inspection requirements but the minimum time savings achieved should be 20% with potential reduction of as much as 70%.

4.1.3. Construction phase

Construction phase is mainly dependent on design and the workmanship on site. Construction phase may be made more sustainable and efficient by improving design techniques, by using reusable construction materials and improving the process of construction. The design process impact on the material quantity.

4.1.4. Operation and Maintenance

Operation and Maintenance phase is important for the service life of building. This phase is very much dominated by the owner. The major energy used in the operation of building is for cooling and heating coupled with the lighting system. Therefore, operation cost can be used by adopting a

combination or renewable and ordinary source of energies. Electricity contribute a lot in operation of buildings. The use can be minimized by putting sensors and automatic system. A lot of natural material are also available in the market for providing glowing lights in nights for that electricity demand is eliminated.

4.1.5. Demolition and reconstruction

During new construction usually the old structure is demolished and disposed off. That debris requires a lot of transportation charges to dispose off. That debris includes bricks, concrete, rocks, lumber, paving, shingle, glass, plastics, steel, aluminum, insulation, asphalt roofing material, electrical materials, plumbing and card boards. That material after demolishing may be sorted out and re-used. That may reduce the project costs through avoided disposal costs and same overall cost when the material is reused.

4.2. Sustainability with respect to construction material

Pre-construction phase with respect to material consists of mainly manufacturing of material (extraction, processing, packaging, shipping and transportation) Construction phase consists of mainly use of material (construction, installation, operation and maintenance. while post construction stage consists of recycling and reuse of material. The rise in cost of construction material is due to higher demand. Rising of demand is again dependent on many factors. In developing countries particularly in south Asia the major factor of this rising depend is high population growth. In figure 1 and 2 the two most import components of construction material (steel and cement) are shown with the price increase every year in Pakistan. This increasing demand of construction material has created a vacuum in the market. Pakistan is set to import 400,000 tonnes of cement to stabilise prices following rising demand. A vessel with nearly 32,000 tonnes of cement is due to arrive from China early this month, while a ship carrying 25,000 tonnes of clinker for making cement is expected from the Philippines soon. Prices in India, Iran, Ukraine and the Philippines range between \$65 and \$70 a tonne, while Chinese cement costs \$65 a tonne. The landed cost of the imported cement comes to about Rs.300-310 per bag after paying various taxes. It would come down to Rs.250 following the rebate.[4] This all can be solved by introducing alternate construction material. Efforts have been made to introduce low cost cement by introducing fly ash (by product of thermal power plants). India is now producing 1,4 m tons of ash and reached a recycling rate of 28%[3]while reinforcement can be replaced by fibre glass. Introduction of fibre glass has been proved successful in flexure member with reasonable good cost. This cost may further be lowered by mass production.

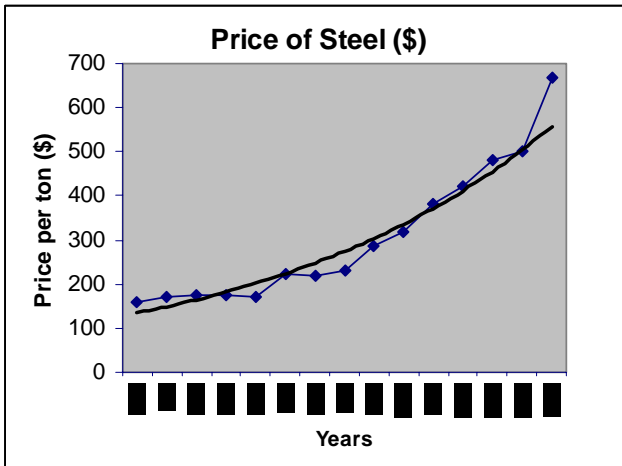


Figure 6. Average annual price of structural steel per tonne

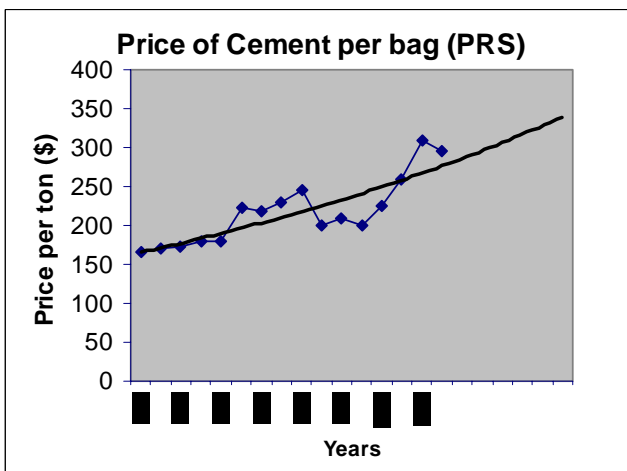


Figure 7. Average annual price of cement per bag



Figure 8. Average annual price of cement per bag

Figure 9. Average annual price of cement per bag

5. Concept of Green Buildings

A green building, also known as a sustainable building, is a structure that is designed, built, renovated, operated, or reused in an ecological and resource-efficient

manner. Green buildings are designed to meet certain objectives such as protecting occupant health; improving employee productivity; using energy, water, and other resources more efficiently; and reducing the overall impact to the environment.

5.1 Economic Benefits of Green Buildings

- A green building may cost more up front, but saves through lower operating costs over the life of the building. The green building approach applies a project life cycle cost analysis for determining the appropriate up-front expenditure. This analytical method calculates costs over the useful life of the asset.
- These and other cost savings can only be fully realized when they are incorporated at the project's conceptual design phase with the assistance of an integrated team of professionals. The integrated systems approach ensures that the building is designed as one system rather than a collection of stand-alone systems.
- Other benefits, such as improving occupant health, comfort, productivity, reducing pollution and landfill waste are not easily quantified. Consequently, they are not adequately considered in cost analysis.



Figure 10. A Model of Sustainable Architecture and Construction

5.2. Elements of Green Buildings

5.2.1. Siting

- A well suited site should selected to take advantage of mass transit.

- Protecting and retaining the existing landscaping and natural features
- Recycled content paving materials, furnishings

5.2.2. Energy Efficiency

- Passive design strategies can dramatically affect building energy performance. These measures include building shape and orientation, passive solar design, and the use of natural lighting
- Develop strategies to provide natural lighting. Studies have shown that it has a positive impact on productivity and well being.
- Installation of high-efficiency lighting systems with advanced lighting controls. Include motion sensors tied to dimmable lighting controls. Task lighting reduces general overhead light levels.
- Use of a properly sized and energy-efficient heat/cooling system in conjunction with a thermally efficient building shell
- Considering alternative energy sources such as photovoltaic and fuel cells. Renewable energy sources provide a great symbol of emerging technologies for the future.
- Computer modeling is an extremely useful tool in optimizing design of electrical and mechanical systems and the building shell.

5.6. Leading towards "Total Satiability" Zero Energy Home (ZEH)

Home that is not only energy efficient, but also produces its own power. Just like a typical home, a **Zero Energy Home (ZEH)**. A ZEH combines state-of-the-art, energy-efficient construction techniques and equipment with renewable energy systems to return as much energy as it takes on an annual basis.



6. Problems associated with sustainable construction in Pakistan

- Resistance in Social and cultural change
- Scarcity of information & knowledge on new technologies
- Force-fed foreign technology and lesser attention towards indigenous /local technologies available
- Limited R & D funding for new built environment technologies

Conclusion

1. Looking a high demand and cost of building material there is urgency in development of alternate, reusable building materials.
2. Sustainable development and construction is not only the responsibility of researchers. It requires concerted action by all stakeholders involved in the creation and use of the built environment.
3. For the researchers to develop this new knowledge, they will need the participation and support of clients, contractors, professionals, governments and regulators. Sustainable construction can therefore take place only if all the necessary elements - both technological and contextual enablers, as well as stakeholders - are developed and work together at local, national, regional and international scale.

Acknowledgments

I wish to acknowledge all those who helped me out in making this report/paper. Particularly, CWHR, PCRET, AKPBS for their support and data. I acknowledge Department of Civil Engineering, University of Engineering and Technology, Lahore for conducting such type of study.

References

- [1] D.M Roodman and N. Lenssen, A Building Revolution: How Ecology and Health Concerns are Transforming Construction, Worldwatch Paper 124, Worldwatch Institute, Washington, DC, March 1995, p. 5
- [2] B.V.Venkatarama Reddy, Current Science Vol. 87 No 7, 10 October, 2004.
- [3] Levin, H. Systematic Evaluation and Assessment of Building Environmental Performance (SEABEP), paper for presentation to "Buildings and Environment", (2001)
- [4] Prof. Ravindra Dhir O.B.E. (Dundee University) UK Quality Ash Association (UKQAA)
- [5] Ashraf M.4/24/2006 10:01:00 AM Source : Business News Onlypunjab.com.
- [8] A.Nani Guide for the design and construction of RCC with FRP bars, University of Missouri, Rolla



Semi-Continuous Composite Connection using Trapezoid Web Profiled (TWP) Steel Beam Section

Anis Saggaff^{1,2*}, Mahmood Md Tahir¹, Shahrin Muhammad¹,
Sariffuddin bin Sa'ad¹, Arizu Sulaiman¹.

¹ Steel Technology Centre, Faculty of Civil Engineering,
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia

² Civil Engineering Department, Faculty of Engineering,
Sriwijaya University, Palembang, Indonesia.

Tel: +60-7-5531627, Fax: +60-7-5576841, E-mail: anissaggaf@yahoo.com

Abstract

The analysis of 'semi-continuous' connections including the moment resistance of 'partial-strength' connections in plastic hinge has been mentioned by the Eurocode 3 for designing steel frames. The use of partial-strength connection in semi-continuous design is complicated compared to simple design. However It gives more economics beams in size and weights. A new design of beam profile section known as Trapezoid Web Profiled (TWP) section is a welded steel girder with the web made corrugated and can be used as an alternative beam for steel frames. TWP girders are also able to be used for composite design. The analysis of composite design can be mentioned theoretically and experimentally, therefore the standardised dimensions can be produces in theoretical analysis. With the same material required, much greater span can be achieved in accordance with the same quantity of steel used. The regular TWP girder usually use the pin joint as a connection. Some joints used by the TWP girder are the fin joint, double angel web cleat, and partial depth flexible end-plate. Flush-end plate is the most probable connection to be used for partial strength connection in composite connection. The savings on weight in the TWP girders will be quite significant. The compression on slab contributes quite significant value to the strength of the beam due to interactions of shear connectors. Composite connection contributes to initial moment capacity that can make the deflection of beam becomes smaller. In addition, using partial strength connections in TWP composite beam will enhance further the moment capacity of the joints. This paper presents the theoretical Analysis of partial strength on TWP composite beams to Universal Beam (UB) using flush end-plate connection. The study is conducted with locally produced steel sections and TWP girders. The theoretical calculation of moment and shear values are based on Eurocode 3 and BS 5950 - 2000 as mentioned in Steel Construction Institute (SCI) publications.

Keywords:

Semi-continuous composite connection; Trapezoid Web Profiled (TWP) steel beam section, TWP composite beam, flush end-plate connection, Universal Column (UC) section.

1. Introduction

A trapezoid web profile plate girder is a built-up section made up of two flanges connected together by a thin corrugated web as shown in Fig 1. The web is corrugated at an angle of 45 degree and welded to the two flanges by using automated machine. The web and the flanges comprised of different steel grade depending on design requirements. TWP section is also classified as a hybrid steel section as two different types of steel grade were used in the development of the section. The steel grade of the flanges is designed for S355 and the steel grade of the web is designed for S275. The flanges is purposely designed for S355 so that the flexural capacity of the beam can be increased and the web is designed for S275 so as to reduce the cost of steel material. The capacity of shear is usually not that critical in the design of the beam. The use of different steel grades in the fabrication of TWP section leads to further economic

contribution to steel frames design besides the use of partial strength connection. The use of thick flanges, thin web and deeper beam for TWP section compared with hot-rolled section of the same steel weight leading to heavier load capacity and greater beam span that can be achieved.

This section can be seen in [figure 1](#). as indicated in the following:

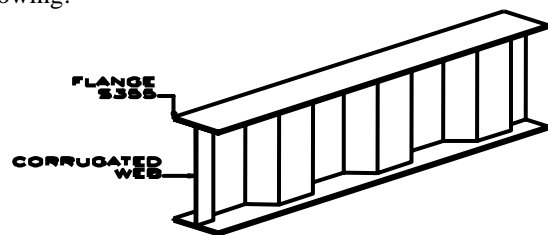


Figure 1. Trapezoid Web Profiled (TWP) beam section to column connection.

The use of composite beam in buildings has known to increase the loading capacity and stiffness of the composite construction. The benefits of composite beam action result in significant savings in steel weight and reduce the depth of the beam. To obtain more economical structural design against the bare steel beams, composite beam was designed by taking the advantage of incorporating the strength of concrete slab by the use of headed studs. The composite action due to the interaction of steel beam and concrete slab with shear connectors increases the load-carrying capacity and stiffness of composite beam. These advantages of composite beam contributed to the dominance of composite beam in the commercial building in steel construction industry. The advantages of composite construction have been further extended with the use of composite connection. This partial strength composite connection enhanced further the stiffness of the connection and increases the weight saving of the design of composite beam.

The termed semi-continuous (partial strength) composite connection is usually associated with a connection having a moment capacity less than the moment capacity of the connected beam. Partial strength composite connection is the term used for connection in the design of semi-continuous construction for multi-storey steel frames by Eurocode 3 where the strength of reinforcement embedded inside the concrete slab was taken into account in improving the moment capacity and stiffness of the connection. In semi-continuous frame the degree of continuity between the beams and columns is greater than that in simple construction design but less than that in continuous construction design. One of connection types of semi-continuous can be seen in [figure 2](#).

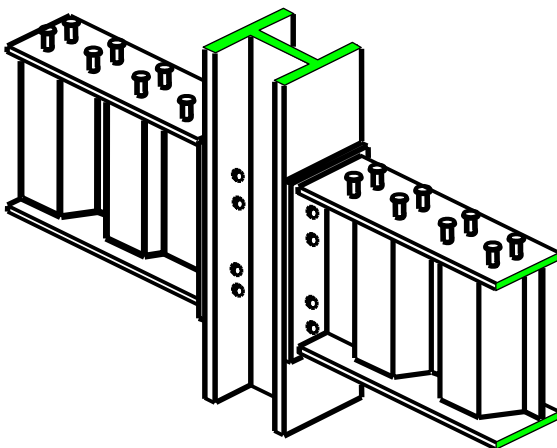


Figure 2. Typical flush-end plate connection for TWP section with studs connected to hot-rolled universal column (UC) section

The degree of continuity in the use of partial strength connection of beam to column can be predicted to produce an economical beam section that representing the section between pin joints and rigid joints. By adopting this approach, studies conducted on the use of partial strength

connection have proven substantial savings in overall steel weight. This is possible as the use of partial strength has contributed to the benefits at both the ultimate and serviceability limit states design. However, the use of partial strength composite connections for Trapezoidal Web Profiled sections has not been established yet. Therefore, this paper intends to establish the standardized tables for partial strength composite connections for TWP sections based on the proposed method by SCI

The design of composite steel frame with partial strength connections is quite fashionable recently. Some advantages of using partial strength design that creates the moment resistance of the connections, therefore the beam size can be reduced. The use of stiffening in the joints can be avoided.

The theoretical aspect of composite connections can be mentioned in two ways; as liner assumption or as plastic assumption. Both of the calculations has different point of view and give quite different result in calculation.

2. Formation Analysis of Trapezoid Web Profiled (TWP) Composite and Connection.

2.1. Concept of TWP Composite.

Trapezoid web profiled (TWP) composite beam is a built-up corrugated web section connected to concrete slabs using studs as shear connectors. This two merged materials can be used as a comprehensive structure that can produce more stronger section in construction.

Composite connections in semi-continuous construction have not been developed yet, more over for TWP composite beams. The studies on the connections as sub-collection of frames have to be calculated in order to mention the behaviour of important aspects on connections, such as the moment resistance, rotational stiffness and rotational capacities of connections in the semi-rigid design. The main objective of the study is to achieve full of meaning on the behaviour of moment connections developing concrete slabs as composite and TWP sections as beams.

A Trapezoid Web Profile Composite beam section can be performed using concrete slab supported with a TWP girder with shears connectors made of studs. The studs are connected to the top flange of the girder by using automated welding machine. The size of the studs varies from 10 mm to 30 mm thick and 90 mm to 140 mm long depending on the need of shear capacity between TWP section and concrete slab. The studs are also designed to resist longitudinal shear forces. The size of the flange can vary from 10 mm to 60 mm thick with its width ranging from 100 mm to 500 mm. The size of flange and the depth of beam will determine the moment capacity of TWP beam. The depth of the TWP beam varies from 200 mm to 1600 mm.

The behaviour of composite connection is more complicated than that of normal steel joints due to the

interaction between the composite floor and the components of the joint, Some research has been done to properly classify the reaction of such connections. This paper mentioned full scale testing used to learn about the behaviour of composite connection.

Laboratory work has proved valuable information on the performance of the composite connections, as well as supplying the obligatory outcome for comparison against analytical work. Analytical work of this research has been mentioned to produce the standard calculation of Composite TWP beam section to column connections.

2.2. Advantages of TWP Composite.

The advantages of TWP section as compared to the conventional plate girder or hot rolled steel section include the following:-

- Utilization of very thin web with minimum thickness of 3mm reduces the weight and the tonnage of the steel.
- Elimination of the need of stiffeners as a result of corrugated shape reduces the fabrication cost.
- The use of high strength steel S355 for flanges and deep beam leads to higher flexural capacity, wider span and less deflection.

Based on the configuration of the structure, TWP beam can offer substantial saving in the steel usage, and in some cases up to 40% as compared to conventional rolled sections. The advantages of use of TWP sections are more significant when there is a need for a column free, long span structural system, such as portal frames for warehouses, girder for bridges, floor and roof beam for high-rise buildings, and portal frame for factory.

3. Composite and Connection Design.

The design philosophy presented in this paper was adopted from 'component approach' described in SCI. The moment capacity of the connection was determined by considering the capacity of each relevant component such as the tensile of the top bolt row and the tensile capacity of the reinforcement bar anchored inside the concrete slab. The composite connection capacity was also checked to meet the requirement of BS 5950:2000 Part1. The moment resistance of the connection was developed by coupling tension force in the reinforcement and the upper bolt with the compression of the beam flange to the column at the lower part of the beam. The lever arm to calculate the moment capacity was established by considering the distance between the components of the tension zone and the compression zone. Composite connection with reinforced concrete and steel decking is shown in figure 3.

3.1. Tension Zone.

The tension zone comprised of three components that govern the magnitude of the tensile force which contributed to the moment capacity of the connection. These three components are listed as the reinforcement bar, the upper row of bolts and the longitudinal shear force.

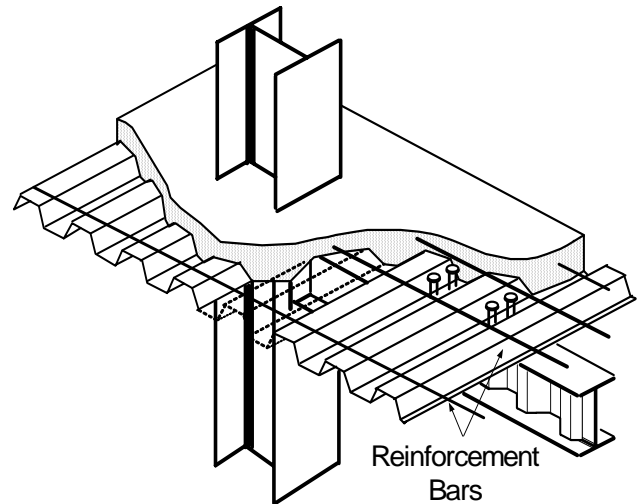


Figure 3. The position of reinforcement around the column and embedded to the slab.

3.1.1. Tensile force of the reinforcement bar.

The position of the reinforcement bar is very important to contribute to the moment capacity of the connection. The distance of the reinforcement bar to the compression flange of the beam will be used in the determination of the moment capacity of the connection. Tests and models have shown that connection rotation capacity increases as the area of reinforcement increases. The minimum area for reinforcement bars is needed to ensure that the connection can undergo sufficient rotation to strain the reinforcement to yield.

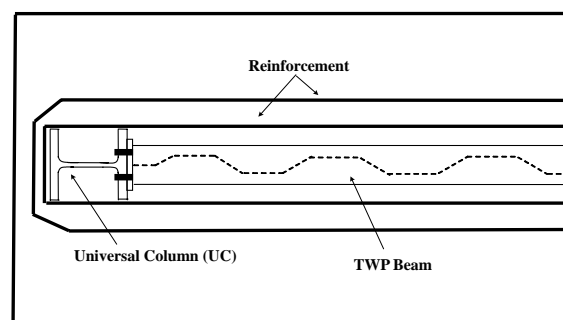


Figure 4. Plain view of reinforcement around the column.

The rebar elongation limit suggested by SCI is up to 10%. The determination of moment capacity is based on the assumption that the reinforcement bar yields. The contribution to moment capacity of any mesh reinforcement

used in the concrete slab should be ignored, because it fails at lower values of elongation than the reinforcement bars. Plain view of reinforcement bars tighten around universal column (UC) on the connection can be seen in figure 4.

3.1.2. Tensile force of the Bolts.

The composite connection is usually use flush end-plate connection where the tension bolt is underneath the upper beam flange. Most of the moment capacity of the connection is developed by the contribution of tension reinforcement bar. However, the contribution of upper bolt row still need to be considered to ensure that the compression zone is not under design which could lead to be premature failure due to non-ductile compression zone. The bolt row furthest from the beam compression flange tends to attract more tension than the lower bolts. The force permitted in the bolt row is based on its potential resistance based on the size of bolt and the thickness of the end-plate, not only the lever arm of the bolt. Details of the formulation and calculation to predict the tension values of the reinforcement and bolts are shown in figure 5 as follows;

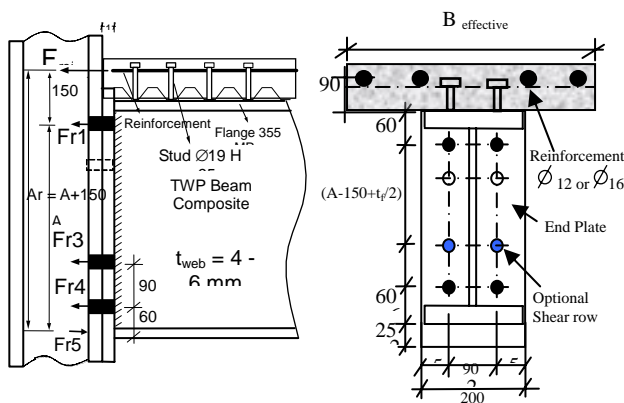


Figure 5. Forces analysis of TWP composite connection.

3.1.3. Longitudinal shear forces.

The formation of full tensile force in the reinforcement depends on the longitudinal shear force being transferred from the beam to the slab by the shear connectors and TWP section as shown in Figure 2. According to the requirement of BS5950; Part 3, full shear connectors should be provided in the region of positive moment. The reinforcement used in the connection should be extended beyond the negative region of the span and anchored sufficiently into the compression region of the slab to satisfy the requirements of BS 8110.

3.2. Compression Zone.

The strength of the compression capacity of the connection relied on the flange and web of the lower beam

and on the resistance of the column web. The failure modes on compression zone are due to crushing or buckling of the column web or on the flange or web of the lower beam. This compression zone however, did not design for the use of stiffener at the web of the column so as to reduce the cost of fabrication. Therefore, the calculated value of moment resistance in the tables developed for TWP section is only for un-stiffened column web.

4. Experimental Works in the Laboratory.

The use of partial strength connection for hot-rolled British sections has well established by SCI. A series of tests at the University of Albertay, Dundee was successfully been carried out to verify the predicted moment and shear capacity with the experimental tests capacities. The results confirmed with the predicted values and the standardized tables for the connection have been published by SCI. However, these tests were carried out for hot-rolled steel section and the connection is a non-composite connection. In this paper, the connection is categorized as composite connection and the sections used were TWP sections. The standard flush end-plate composite connections tables for TWP sections were developed in accordance to Eurocode 3 and SCI procedures.

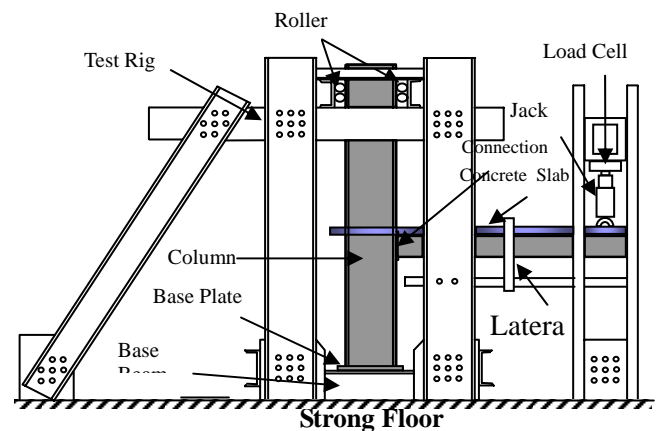


Figure 6. Test rig for Experiments in the Laboratory

In the paper it can be seen that the best validation of the results presented in the tables is by comparing the predicted results from theoretical calculations with extensive experimental tests results. However, due to higher cost to conduct an extensive full scale testing, four tests have been carried out to validate the presented standard connection tables for TWP section. The test rig of experiment draw is shown in figure 6 and the specimens properties are listed in Table 1, from which it can be seen that the parameters were varied in systematic manner. In the experiment, four number of 16mm diameter reinforcement bar were used and the thickness of the slab is 125mm with the concrete grade of G-30.

Table 1. Test specimens with various parameters.

Speci men.	Bolt Rows	Bolts	End Plates			Size of TWP Beam	Size of Column
			W	D	T		
C5	1	M20 8.8	200	440	12	400x140 x39.7	305x305 x118
C6	1	M24 8.8	250	540	15	500x180 x61.9	305x305 x118
C7	2	M20 8.8	200	490	12	450x160 x50.2	305x305 x118
C8	2	M24 8.8	250	640	15	600x200 x73.3	305x305 x118

5. Test Procedures.

Test specimens were set-up by connecting a 3 m long column with a 1.5 m long beam as shown in figure 7. A metal decking with 1.5 m width which acts as a permanent formwork for the slab was attached to the top flange of the column by a pair of shear stud on each though. The shear studs were measured as 19mm diameter and 95mm height was used. The thickness of the slab was taken as 125mm thick with concrete strength of grade 30. Four reinforcement of size of 12mm or 16mm diameter was installed around the column and embedded to the slab as shown in figure. 3 and figure 4. The top and bottom part of the column was restrained from any movement. Point load was applied at 1.2m from the centre of column flange.



Figure 7. Experimental works during the test in the Laboratory.

Load was applied through an automatic operated hydraulic jack and monitored with a pre-calibrated 100 tonnes capacity load cell. The data logger system was set-up to read displacements from inclinometer in millimeters and load in kN. A small load was applied and then removed, to check the performance of the rig. Significant load was then applied, sufficient to cause extensive inelastic deformation of the connection. To determine the complete response, each connection was later subjected to unloading, followed by

reverse loading. The following load sequence was generally used. An increment of 5 kN was applied to the specimen. The readings for loads, displacement, and rotation were recorded after two minutes had elapsed. This time elapse allowed the specimen to reach an equilibrium state. The incremental load procedure was then repeated until there was a significant increase in deformation. The loading on the specimen was then controlled by deflection increments of 3mm. The test was continued until failure, when large deformation or the load decreases significantly. The response of a joint in these phases may govern the buckling behavior of the connected column. A graph of rotation of the connection was plotted against moment to predict the moment resistance of the connection.

6. Prediction of Moment Resistance (MR) and Discussions.

The prediction of moment resistance and the lost in stiffness is very much dependent on the stiffness of the connected members, types of joints, and orientation of the column axis. Beam-to-column connections generally have non-linear moment-rotation curves. Initially, the connections have a stiff initial response which is then followed by a second phase of much reduced stiffness. This second phase is due to in-elastic deformation of the connections' components or those of members of the frame in the immediate vicinity of the joint. These deformations need to be accounted for because they contribute substantially to the frame displacements and may affect significantly the internal force distribution. The structural analysis needs to account for this non-linearity of joint response to predict accurately both stiffness and resistance for a semi-continuous frame in case the joint behaviour exhibits a form of material non-linearity. The examples of shape of the experimental results for the M-Φ curve are shown in figure 8 to figure 11. The experimental values of moment resistance MR listed in Table 2 was determined by estimating when a "knee" formed in each of the M-Φ curves plotted Figure 8 to Figure 11.

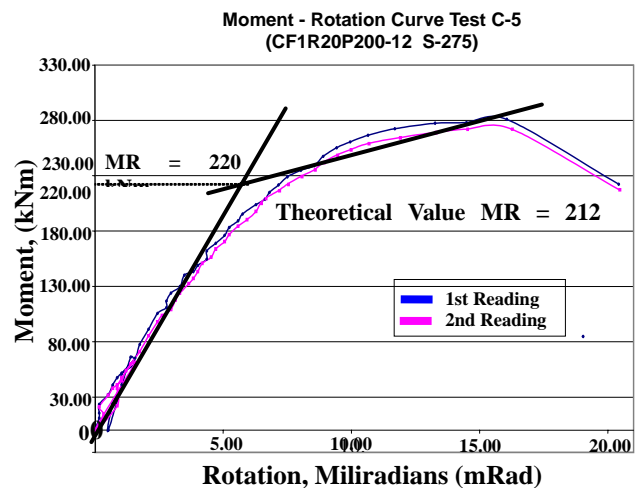


Figure 8. Moment Rotation Curve Test Result for C-5

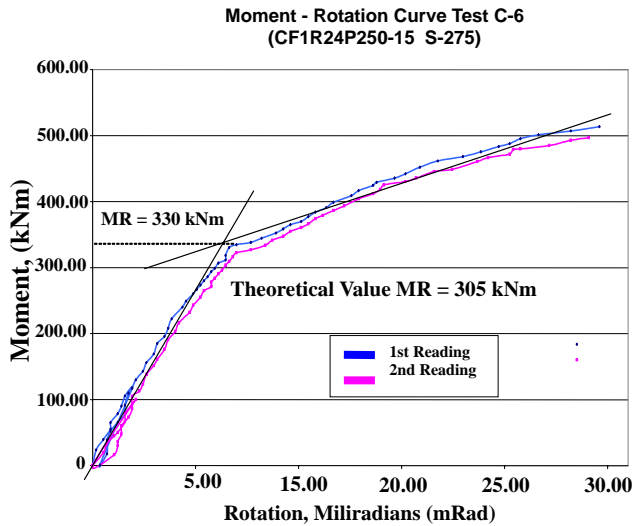


Figure 9. Moment Rotation Curve Test Result for C-6

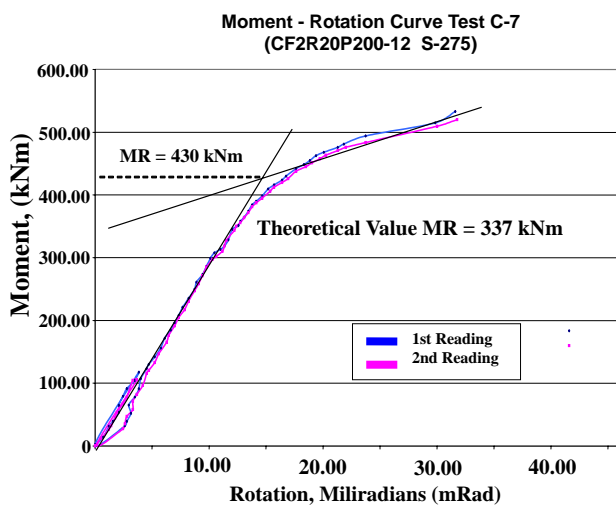


Figure 10. Moment Rotation Curve Test Result for C-7

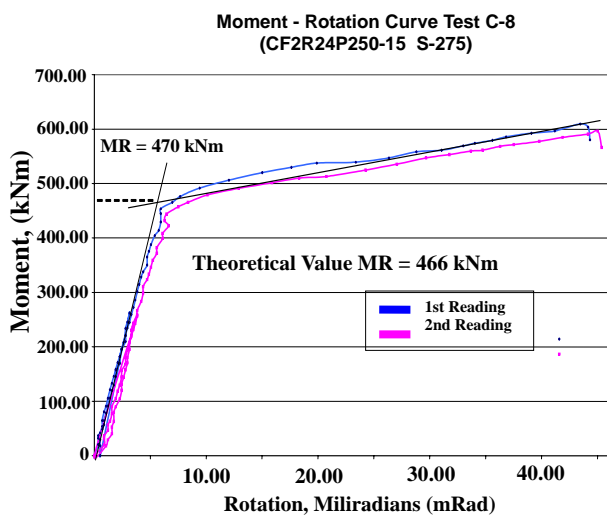


Figure 11. Moment Rotation Curve Test Result for C-8

By adopting this technique, the experimental values of MR for the overall joint for the tests were established. For

curves which do not clearly show a linear region, an assumed straight line was drawn parallel to the unloading region traced from the exact $M-\Phi$ curves.

Table 2: Theoretical and experimental values of moment resistance for each specimen.

Specimen	Moment Resistance, MR (kNm)		Ratio of Theoretical vs. Experimental values
	Theoretical values	Experimental values	
C-5	212.0	220.0	1.04
C-6	305.0	330.0	1.08
C-7	337.0	430.0	1.23
C-8	466.0	470.0	1.01

The results of the moment resistance of composite connection were also compared with the moment resistance of partial strength connection of extended end-plate connection. The moment resistance of the extended end-plate connection was established by applying the component method proposed by Steel Construction Institute. The comparison of the results was shown in Table 3. From the results, the use of proposed composite connection for TWP section has shown an increment up to 72.4% of the moment resistance of the connection.

Table 3: Comparison of moment capacities for composite and non-composite connections

Specimen	MR (kNm)		% Difference
	Composite	Non-composite (extended end-plate)	
C-5	212.0	123.0	72.4
C-6	305.0	181.0	67.0
C-7	337.0	235.0	43.4
C-8	466.0	432.0	7.9

The graphs showed that the connections behaved linearly in the first stage followed by non-linear behaviour and gradually losing the stiffness with the increase in rotation. The overall results showed that the experimental values of moment resistance were greater than the theoretical values with the ratio ranged in between 1.01 to 1.23 as shown in Table 1.

7. Conclusions.

This study concluded that it is possible to determine the moment resistance of composite flush end plate connections connected to a column flange by adopting the method proposed by SCI, even for different geometric parameters such as TWP section. The capacities of the connection depends on the geometrical aspects of the connection such as the size of bolt, number of bolt, size of end-plate, thickness of end-plate, size of beam, size of column and the size and number of reinforcement.

The increment of moment capacity of the connection can be concluded as follows:-

- The increase in the number of bolt row from one row to two rows has contributed to an increase in the moment capacity up to 23% which is quite significant.
- The use of reinforcement in composite connection has increased the moment resistance of the connection up to 73% compare with the extended end-plate connection for non-composite connection.
- The ductility of the connection in composite connection is suitable for the design in semi-continuous construction as the ductility of the connection is more than 20mrad as proposed by SCI.

Acknowledgements.

This study is part of a research towards a PhD by one of the authors. The authors would like to acknowledge special thanks and gratitude to CIDB for funding this project under Vot 73049. Special thanks is also addressed to Steel Technology Centre (STC) and all staff of Structure Laboratory, Universiti Teknologi Malaysia for some efforts conducting with this research that could bring this experiment awarded gold medal on Malaysian Technology Expo (MTE) 2006 and International Invention Exhibition 2006 in Geneva.

References

- [1] Abdalla, K. M. and Chen, W. F. (1995). Expanded Database of Semi-Rigid Steel Connections. Computer And Structures, Vol. 56, No. 4, pp 553-564.
- [2] Bose, B. (1993). Tests to verify the performance of standard ductile connections. Dundee Institute of Technology.
- [3] BRITISH STANDARDS INSTITUTION. (1990). BS 5950 (1990): Structural use of steelwork in building Part 3: Design in composite construction Section 3.1: 1990: Code of practice for design of simple and continuous construction: hot rolled sections BSI.
- [4] BRITISH STANDARDS INSTITUTION. (1994). DD-ENV 1994-1-1: 1994 Eurocode 4: Design of composites steel and concrete structures Part 1.1 General rules for buildings EC 1994.
- [5] BRITISH STANDARDS INSTITUTION. (1997). BS 8110: Structural use of concrete. Part 1: Code of practice for design and construction.
- [6] BRITISH STANDARDS INSTITUTION. (2000). BS 5950-1. Structural Use of Steelwork in Building Part 1: Code of Practice for Design – Rolled and Welded Sections. British Standards Institution, London
- [7] Chen, W. F. et .al. (1993). Semi-rigid Connections in Steel Frames. Council on tall Buildings and Urban Habitat, Committee 43, Mc Graw-Hill, New York.
- [8] Chen, W. F. and Kishi, N. (1989). Semi-rigid Steel Beam-to-Column Connections: Database and Modelling. Journal of Structural Engineering. Vol. 115, No. 1, pp 105-119.
- [9] Eurocode 3. (1992). Design of Steel Structures: ENV 1993-1-1: Part 1.1: General Rules and Rules for Buildings. CEN, Brussels.
- [10] EUROPEAN COMMISSION, COST C1. (1994). Semi-rigid behaviour of civil engineering structural connections, composite steel-concrete joints in braced frames for buildings EC 1994.
- [11] Couchman, G. H. (1997). Design of Semi-continuous Braced Frames, Steel Construction Institute Publication 183, Silwood Park, Ascot, Berkshire SL5 7QN, U.K
- [12] Hussein, Wa' il Q. (2001). Design Guide for Steel Plate Girder with Corrugated Webs (TWP). Presentation in Design of Steel Structure Short Course, TWP Sdn Bhd.
- [13] Jaspart, J. P. (2000). General Report: Session on Connections. Journal of Constructional Steel Research. Vol. 55, pp 69-89.
- [14] Luo, R. (1995). Load Carrying Capacity of Steel Girders and Panels with Thin-Walled Trapezoidally Corrugated Webs. Compilation of Papers. Chalmers University of Technology, Sweden.
- [15] Tahir, M.M. (1995). Structural and Economic Aspects of The Use of Semi-Rigid Joints in Steel Frames. PhD Thesis. University of Warwick, United Kingdom.
- [16] Tahir, M.M. (2003). Design of Semi-Continuous Construction for Multi-Storey Braced Steel Frames Using TWP sections. Technical report, Steel Technology Centre, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia.
- [17] Osman, M. H. (2001). Performance Test and Research on Trapezoid Web Profile. Presentation in Design of Steel Structure Short Course, Universiti Teknologi Malaysia.
- [18] Saggaff, A. (2006). Economic aspects of composite partial-strength connection in semi-continuous construction for multi-storey braced steel frame using TWP sections, Civil Engineering Faculty, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia.
- [19] Steel Construction Institute and British Constructional Steelwork Association. (1992). Joints in Simple Construction. Volume 2: Practical Applications. First Edition, Silwood Park, Ascot, Berks SL 7QN, London.
- [20] Steel Construction Institute and British Constructional Steelwork Association. (1994). Joints in Simple Construction. Volume 1: Design Methods. Second Edition, Silwood Park, Ascot, Berks SL 7QN, London.
- [21] Steel Construction Institute and British Constructional Steelwork Association. (1996). Joints in Steel Construction. Volume 1: Moment Connections. London.
- [22] Steel Construction Institute and British Constructional Steelwork Association, SCI P213 (1998), Joints in Steel Construction Composite Connections, Silwood Park, Ascot, Berks SL 7QN, London.

Effect of Aggregate Gradation to Workability of Hot Asphalt Mixture

Sri Widodo , Ika Setianingsih

Department of Civil Engineering,
Muhammadiyah University of Surakarta Indonesia
Tel: +62-271-717417, Fax +62-271-715448, E-mail : sri_widodo@ums.ac.id

Abstract

The objectives of this research are to examine workability of hot asphalt mixture, which are consisting of Asphaltic Concrete (AC), Hot Rolled Sheet (HRS) and Asphalt Treated Based (ATB). The examination is using Marshall testing machine at 5 aggregate gradations. The 5 aggregate gradation are at lower specification, middle specification, upper specification, between lower and middle specification and between upper and middle specification.

Testings were executed at optimum asphalt content and compaction temperature 140°C. Optimum asphalt content was found by using middle specification of aggregate gradation. Testings were done by Marshall testing machine as substitute Gyratory testing machine with equivalent factor 2x1 Marshall blow equal with 1.6 revolution Gyratory testing machine. Measurement of porosity was made at number blow of Marshall 2x5, 2x10, 2x20, 2x40, 2x60 and 2x75. From the equation between log number of revolution and porosity, Workability Index (WI) every hot asphalt mixture can be calculated.

Research results shows that all hot asphalt mixture are not including the workable criteria, because the WI are less than 6. However, from the three types of hot asphalt mixture, ATB proved the best workability with average WI = 2,98, followed by HRS with average WI = 2,80 and than AC with average WI = 2,64.

Key word: gradation, asphalt mixture, porosity, workability

1. Introduction

Most of all surface course of road in Indonesia use the hot asphalt mixture, because this road surface course is assumed to give the comfort to road user as well as the initially construction cost which is relative cheaper in comparison with surface course using cement concrete or known as rigid pavement. The hot asphalt mixture has been used much in Indonesia, for example Asphaltic Concrete (AC) and Asphalt Treated Based (ATB) which are having the structural characteristic. Besides both types of the hot asphalt mixture are also known as the hot asphalt mixture that the main function are not as structural course, but as wearing course. For example Hot Rolled Sheet (HRS) and Hot Rolled Sand Sheet (HRSS), see [1].

During construction, failure of work can be caused by the difficulty of materials to be laid and compacted or the materials have the bad workability. This matter is caused by a hot asphalt mixture materials including the thermoplastic materials, of which the plasticity depends on the temperature, see [3]. Materials having low plasticity will be difficult to be compacted. However, if the materials are too plastic, the sufficient stability is also difficult to reach at the moment of compaction work. Beside problem of temperature, workability mixture is also depends on composition between aggregate and asphalt in hot asphalt mixture. Aggregate gradation will also influence the workability of mixture. Aggregate with dense gradation tend to have stiff character, but aggregate with open gradation is likely to have the character more flexible.

Research concerning workability of hot asphalt mixture can be conducted to use the Gyratory compaction

machine, see [4]. However, because appliance compaction test by Gyratory machine is not yet recognized in Indonesia, the research will be done by approach using the Marshall compaction machine, which more well known in Indonesian both by academician and practitioner.

Problems to be researched are as follows:

- How far is the influence of gradation of aggregate mixture on workability of hot asphalt mixture ?
- Which is the best workability between Asphaltic Concrete, Hot Rolled Sheet and Asphalt Treated Base ?

The aims of this research are to measure the workability of Asphaltic Concrete, Hot Rolled Sheet and Asphalt Treated Base at various aggregate gradation. Aggregate gradation used are still in specification boundary determined in general specification of Rehabilitation and Maintenance Project of Road and Bridge, see [1]. Results of this research are expected to give the guidance of preparation of workable hot asphalt mixtures to all road-works construction in Indonesia for the owner, contractor and supervision consultant.

Method developed by Cabrera (see [5]) to measure the workability of hot asphalt mixture by using appliance of Gyratory Testing Machine (GTM) consist of the following steps:

- Monitoring the reduction of specimen height during compaction process
- Using the specimen height registered to calculate the specimen volume and its porosity at a number of revolutions, using the following formula:

$$V_i = \frac{10^2}{4} \pi x h_i \quad (1)$$

$$D_i = \frac{W_a}{V_i} \quad (2)$$

Where:

V_i = the Volume of a specimen at i revolutions (cm³),

h_i = Height of specimen at i revolutions (cm),

D_i = Density at i revolutions (gram / cm³),

W_a = Weight of the specimen in air (gram)

The Specific Gravity (SG) of each specimen is obtained by using the following formula:

$$SG = \frac{100}{\frac{P_{wa}}{SG_a} + \frac{P_{ws}}{SG_s} + \frac{P_{wf}}{SG_f} + \frac{P_{wb}}{SG_b}} \quad (3)$$

and the porosity at i revolutions :

$$P_i = 1 - \frac{D_i}{SG} \times 100 \quad (4)$$

Where:

P_i = total porosity at i revolutions (%),

SG = Specific gravity of specimen,

P_w = percentage weight in mixture, a = coarse aggregate, s = sand, f = filler, b = bitumen.

Based on calculation result a graph relating P_i to the \log_{10} [number of revolution (i)] is plotted. The graph should approximate a linear with the equation:

$$P_i = A - B \log_{10}(i) \quad (5)$$

where :

A = intercept with the y axis,

B = slope of the line,

i = number of revolutions.

The Workability expressed by Workability Index (WI) is defined with the formula:

$$WI = \frac{I}{A} \times 100 \quad (6)$$

Experience with Hot Rolled Asphalt of gap graded indicated that the mixtures by $WI < 6$ were difficult to be handled and compacted. According to Zocrob (see [5]) the number of compactive revolution is depending on the expected traffic. For the medium compaction, 80 revolution of Gyratory compactor are applied which is equivalent to 50 blows of the Marshall Hammer. While for the heavy compaction, 120 revolutions Gyratory compactor is applied which is equivalent to 75 blows of Marshall Hammer. It is concluded that 1 blow of Marshall Hammer is equivalent to 1,6 revolution of Gyratory compactor.

2. RESEARCH METHOD

Research was conducted with the following steps:

- Testing of raw materials, covering the testing of specific gravity of the coarse aggregate, fine aggregate, and bitumen materials,
- Conducting Marshall testing to determine the optimum asphalt content for every mixture of AC, HRS and ATB,
- Testing of mixture Porosity to various sum of blows by means of Marshall machine to obtain the relation graph between energy of compaction and porosity of mixtures AC, HRS and ATB. Five kinds of the blows of Marshall Hammer are 5, 10, 20, 40, 60 and 75 times. The Testing of porosity mixtures were applied to 5 aggregate gradation. Five kinds of the aggregate gradation included the lower limit, middle, upper limit, between lower and middle and between upper and middle.

3. RESEARCH RESULT AND DISCUSSION

3.1. Variation of Gradation

Variation of gradation in research of workability asphalt mixture covered the lower limit gradation then called as gradation 1, middle gradation as gradation 3 upper limit as gradation 5, between gradation 1 and gradation 3 as gradation 2 and between gradation 3 and gradation 5 as gradation 4.

For the asphalt mixture of AC, the five kinds of variation of gradation are shown in Figure 1. The five variation of gradation of HRS asphalt mixture are shown in Figure 2. While five variation of gradation of ATB asphalt mixture are shown in Figure 3. All specimens are compacted at temperature 140°C.

3.2. Specific gravity and density of mixture

Results of calculation of specific gravity to the each of gradation type are presented in Table 1. Density of each mixture type will depend on number of blow to the each gradation type. Density is the ratio of specimen weight on the air to the specimen volume. Results of density calculation to each gradation type are presented in Table 2

Asphalt mixture will have the good workability if the asphalt mixture quickly reach the maximum density when it is compacted by compaction machine. The mixture workability is considered good if, when compaction energy increase the changing of porosity is not too big.

3.3. Porosity of asphalt mixture

By using the equivalent between 2x1 blow of Marshall hammer with 1,6 revolution of Gyratory compactor, the porosity of hot asphalt mixtures at various revolution of Gyratory compactor are as shown in Table 3. Based on relation between porosity and logarithm of amount of Gyratory compactor revolution, it can be determined intercept (A)

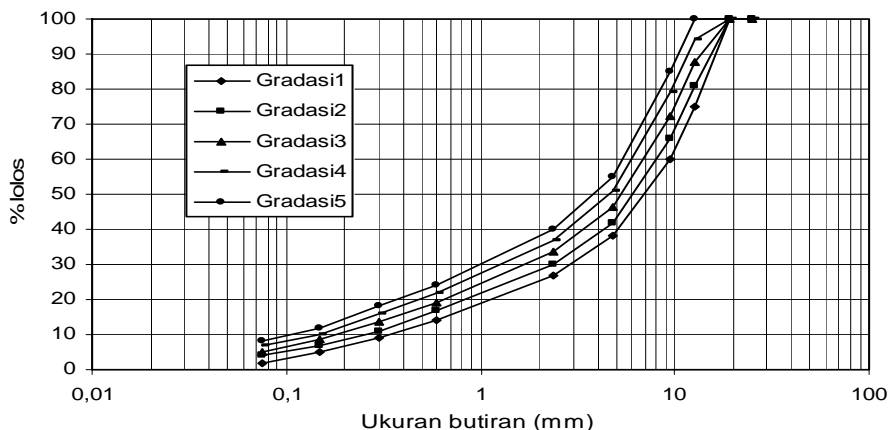


Figure 1: Variation of gradation of AC mixtures

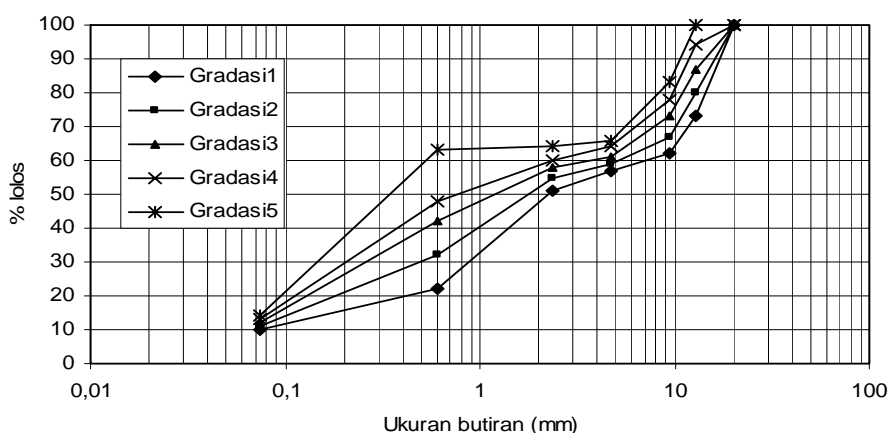


Figure 2: Variation of gradation of HRS mixtures

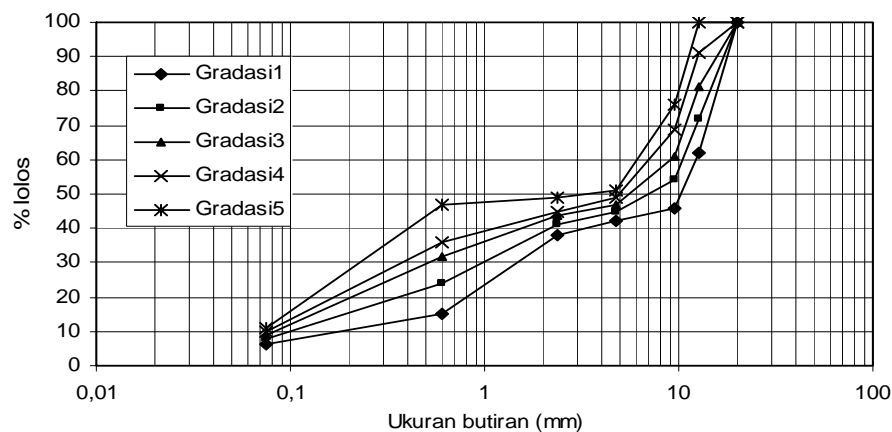


Figure 3: Variation of gradation of ATB mixtures

Table 1: Specific gravity of mixture to the each of gradation type

Gradation type	Type of mixture		
	AC	HRS	ATB
1	2,396	2,349	2,360
2	2,395	2,349	2,359
3	2,393	2,348	2,358
4	2,391	2,347	2,357
5	2,389	2,346	2,357

Table 2 : Calculation result of density (gr/ cm³) of hot asphalt mixture

Type of mixtures	Type of gradation	Number of blow to Marshall hammer					
		2x5	2x10	2x20	2x40	2x60	2x75
AC	1	1,92	2,00	2,09	2,19	2,29	2,33
	2	1,85	1,93	2,00	2,12	2,23	2,31
	3	1,85	1,91	2,02	2,15	2,30	2,35
	4	1,87	1,95	2,08	2,17	2,24	2,30
	5	1,78	1,86	1,95	2,06	2,15	2,24
HRS	1	1,84	1,93	2,08	2,14	2,21	2,32
	2	1,84	1,93	2,06	2,15	2,23	2,29
	3	1,81	1,87	1,94	2,01	2,09	2,13
	4	1,82	1,89	1,95	2,08	2,16	2,22
	5	1,76	1,86	1,94	1,99	2,04	2,09
ATB	1	1,92	2,00	2,10	2,21	2,26	2,32
	2	1,82	1,90	2,02	2,12	2,26	2,32
	3	1,92	2,01	2,11	2,22	2,30	2,34
	4	1,93	2,03	2,11	2,16	2,22	2,28
	5	1,80	1,86	1,98	2,06	2,11	2,17

Tables 3.Result of porosity calculation (%) and intercept (A) hot asphalt mixture

Type of asphalt mixture	Type of gradation	Porosity (%) at equivalent number revolution of Gyrotory compactor						Intercept A
		8	16	32	64	96	120	
AC	1	20	17	13	9	5	3	33,810
	2	23	20	17	12	7	4	38,638
	3	23	20	15	10	4	2	40,843
	4	22	18	13	9	6	4	35,878
	5	26	22	18	14	10	6	41,113
HRS	1	22	18	11	9	6	1	37,044
	2	22	18	12	9	5	3	36,598
	3	23	20	17	14	11	9	33,766
	4	23	20	17	11	8	6	37,251
	5	25	21	17	15	13	11	34,673
ATB	1	19	15	11	6	4	2	32,158
	2	23	19	14	10	4	2	39,848
	3	19	15	11	6	3	1	33,147
	4	18	14	11	8	6	3	28,511
	5	24	21	16	13	11	8	36,129

3.4.Workability Index (WI) of asphalt mixtures

Workability Indexes of hot asphalt mixture is

calculated with the formula 6 : $WI = \frac{I}{A} \times 100$.

Calculation results of WI at various of gradation at optimum asphalt content are shown in Table 4. Increasing the Workability Index of hot asphalt mixture will make the mixture progressively easy to be done and compacted. The limit of categorized of hot asphalt mixture is easy to be done if the WI is greater than 6.

Table 4. Calculation result of Workability Index (WI)

Type of gradation	Type of asphalt mixtures		
	AC	HRS	ATB
1	3,0	2,7	3,1
2	2,6	2,7	2,5

3	2,4	3,0	3,0
4	2,8	2,7	3,5
5	2,4	2,9	2,8
Average	2,64	2,80	2,98

From Tables 4, it can be seen that they are no significant differences value WI to various type of hot asphalt mixture by variation of gradation. All of the hot asphalt mixture type can be considered as workable mixture. However, from testing results of 3 types of hot asphalt mixture, it is seen that ATB mixture with WI = 2.98 is most workable, then followed by HRS mixture with WI = 2.80 and AC mixture with WI = 2.64.

CONCLUSIONS AND RECOMENDATIONS

From the result of research of workability by Marshall testing machine to hot asphalt mixture of AC, HRS and ATB, the following conclusions are offered:

1. Aggregate gradation does not give the significant influence to mixture workability, although all type of examined mixture cannot be categorized to have the good workability.
2. Although all mixtures are not considered to have the good workability, the mixture ATB represent the most workable mixture with WI of 2,98, followed by HRS mixture with WI of 2,80 and AC mixture AC with WI of 2,64.

From the result of research, the following recommendations are offered:

1. Considering research result obtained, it shows the values of Workability Index which are not far different for various of hot asphalt mixture type, it is recommended to reexamination the research by using Gyrotory compaction machine.
2. It is recommended to perform further research of workability with the variation of compaction temperature, because variation of this compaction

temperature also represent one of the problem possibly occurring in field.

REFERENCES

- [1] -----, 1998, *Spesifikasi Umum Proyek Rehabilitasi/Pemeliharaan Jalan dan Jembatan Propinsi DIY*, Direktorat Bina Teknik Direktorat Jenderal Bina Marga Departemen Pekerjaan Umum.
- [2] -----, 1974, *Standard Specifications for Transportation Materials and Methods of Sampling and Testing Part II Method of Sampling and Testing*, The American Association of State Highway and Transportation Officials.
- [3] -----, 1990, *The Shell Bitumen Handbook*, Shell Bitumen UK
- [4] -----, 1997, *Superpave Performance Graded Asphalt Binder Specification and Testing*, Asphalt Institute Lexington Kentucky
- [5] Cabrera, J.G. and Zocrob,S.E.,1999, *Performance and Durability of Bituminous Materials and Hydraulic Stabilised Composites*,Proceeding of the Third European Symposium, Aedificatio Publishers Zurich.

The Effect of Road Section Characteristics on Safety Belt Use Rate in Jakarta

Leksmono Suryo Putranto¹, Budi Fitrianto¹, Achmad Khusaini¹

Department of Civil Engineering

*Tarumanagara University, Jl. Let. Jen. S. Parman No. 1, Jakarta 11440, Indonesia
Tel: +62-21-5672548, Fax: +62-21-5663277, E-mail: lexy@tarumanagara.ac.id*

Abstract

The enforcement of safety belt use in Indonesia has been started from 5 May 2004. Previous research regarding safety belt use in Jakarta was carried out by Putranto (2005) through interview to 50 respondents which were drawn randomly using list of Jakarta's fixed phone customer. It was found that 78% of the respondents always wear safety belt and 22% often or sometimes wear safety belt. 64% of those who don't always wear safety belt stated that road location is an affecting factor, whilst 27% of them are affected by time of the day. Therefore this paper will discuss the result of observation on safety belt use rate in different road sections in Jakarta with various characteristics. Observations were carried out in 7 road sections with various number of lanes, road widths, and V/Cs (volume-capacity ratios). Pearson correlation analysis was carried out between safety belt use rates and number of lanes, road widths, and V/Cs ratios. The basic assumption was that number of lanes, road widths and V/Cs represent road activity level. Higher level activity of the road was assumed to be affecting vehicle users awareness of being monitored by the law enforcement officers. However the results show that number of lanes and road widths were not suitable to predict safety belt use rate of the driver and the front seat passenger. For two lanes two ways road, a significant and high positive correlation was found between safety belt use rates and V/Cs. On the contrary for four lanes and six lanes two ways divided road, a significant and negative correlation was found between safety belt use rates and V/Cs. This suggests that in wider road, driver awareness of wearing seat belt might be more affected by the speed (an inverse of V/C) due to possibility to drive in higher speed.

Keywords:

Safety Belt Use Rates, Road Section Characteristics, Jakarta

1. Introduction

The enforcement of safety belt use in Indonesia has been started from 5 May 2004. This is to implement article 23 paragraph 1e of Law No.14/1992 about Traffic and Road Transport, i.e. that whilst driving a four or more wheels vehicle in a road, a driver is required to use seat belt. On the paragraph 2 of the same article, the front seat passenger is also required to wear the seat belt.

In the USA, there are two type of safety belt law, i.e. primary and secondary laws. Each of the state has its own law. In a primary law, a police is allowed to stop a motorist solely for being unbelted [1], whilst in a secondary law, a seat belt use enforcement is only permitted after a stop for another violation [2]. Safety belt use in Indonesia is enforced based on primary law. In 1984, the safety belt use rate in the USA was about 20%. Five years later, in states where primary law, secondary law and no law were implemented, the safety belt use rates were about 70%, 50% and 35% respectively [3]. This indicates that although primary law was the most effective enforcement to increase the safety belt use rate, in the states in which no safety belt law were implemented, there was an increase in the safety belt use rate. This might be affected by other promotion measures such as combination of public information, education and law enforcement [4] or incentive [5].

Several demographic and socioeconomic factors summarized by [6] that were indicated by previous research to be influencing the safety belt use rate, i.e. age, gender, education, socioeconomic level, race, health condition, job satisfaction, life satisfaction, behaviour and car age. [7] stated that there was no significant relationship between the perceived usefulness of safety belt and the safety belt use rate. According to Hong [8] and [9] parents that always use belts and enforce use by their children will probably raise adult belt users.

Previous research regarding safety belt use in Jakarta was carried out by [10] through interview to 50 respondents which were drawn randomly using list of Jakarta's fixed phone customer. It was found that 78% of the respondents always wear safety belt and 22% often or sometimes wear safety belt. 64% of those who don't always wear safety belt stated that road location is an affecting factor, whilst 27% of them are affected by time of the day. Therefore this paper will discuss the result of observation on safety belt use rate in different road sections in Jakarta with various characteristics.

2. Approach and Methods

Observations were carried out in 7 road sections with various number of lanes, road widths as listed in Table 1.

Table 1. The Characteristics of the Observed Links

Road Link	Road Type*	Road Type Code	Road Width (m)
Otista	6/2 D	3	26.2
TMII	6/2 D	3	22.6
Halim PK	4/2 D	2	12.4
Raya Bogor	4/2 D	2	13.6
Cawang Baru	2/2 UD	1	8.9
Condet	2/2 UD	1	5.8
Pinang Ranti	2/2 UD	1	6.8

* 6/2 D : 6 lanes-2 ways divided
 4/2 D : 4 lanes-2 ways divided
 2/2 UD : 2 lanes-2ways undivided

There were 3 main data collection activities conducted, i.e. road section characteristics survey, classified traffic count and safety belt use survey.

During the road section characteristics survey, the width of the traffic lane, median, shoulder/ kerb were measured. The side friction level on each section was also identified (this information is required for road capacity analysis).

Classified traffic count was carried out to get the 2-way traffic flow of each link. The vehicles were classified into:

- A. Passenger car (private)
- B. Taxi
- C. Paratransit
- D. Pick-up/Box
- E. Public Bus
- F. Non-public bus
- G. Truck
- H. Motorcycle
- I. Unmotorized vehicle

Later, for volume to capacity ratio(V/C) analysis using Indonesian Highway Capacity Manual (IHCM) 1997, the observed vehicles were re-classified into light vehicle (A,B,C,D), heavy vehicle (E,F,G), motorcycle (H) and unmotorized vehicle (I).

Safety belt use survey was carried out to find out the difference of safety belt use rate in road links with different number of lane, width and V/C. In this survey, vehicles were classified into 7 vehicle types from A to G if the classification for the traffic count is used. Only the vehicles in the inner lanes of the busiest directions were observed.

Classified traffic count and safety belt use rate were carried out in a morning peak hour period (07.00-09.00). Sub-period of survey was 15 minutes (every 15 minutes, the count was grouped). According to IHCM 1997, the pcu values are as presented in Tables 2 and 3.

Capacity analysis was carried out according IHCM 1997. Urban road capacity can be formulated as equation (1) and the description of the equation is listed in Table 4.

Table 2. Pcu Values for Undivided Urban Roads

Road Type*	2 Ways Total Traffic Flow (vehicles/hour)	Heavy Vehicle	pcu	
			Motorcycle	
			Lebar Jalur, W _c (m)	
			<6	>6
2/2 UD	<1800	1,30	0,50	0,40
	≥1800	1,20	0,35	0,25
4/2 UD	<3700	1,30	0,40	
	≥3700	1,20	0,25	

*2/2 UD : 2 lanes-2 ways undivided
 4/2 UD : 4 lanes-2 ways undivided

Table 3. Pcu Values for Divided Urban Roads

Road Type*	Traffic Flow per Lane (vehicles/hour)	smp	
		KB	SM
2/1 and 4/2 D	<1050	1,30	0,40
	≥1050	1,20	0,25
3/1 and 6/2 D	<1100	1,30	0,40
	≥1100	1,20	0,25

*2/1 : 2 lanes-1 way
 3/1 : 3 lanes-1 way
 4/2 D : 4 lanes-2 ways divided
 6/2 D : 4 lanes-2 ways divided

$$C = C_0 FC_w FC_{SP} FC_{SF} FC_{CS} \tag{1}$$

Table 4. Description of Equation (1)

Variable	Description	A Function of
C	capacity (pcu/hour)	
C ₀	base capacity (pcu/hour)	road type
FC _w	lane width factor	road type, effective width
FC _{SP}	directional split factor	road type, directional split
FC _{SF}	side friction factor	road type, side friction class, shoulder/ kerb width
FC _{CS}	city size factor	population

A correlation analysis was carried out between safety belt use rates and number of lanes, road widths, and V/Cs ratios. The basic assumption was that number of lanes, road widths and V/Cs represent road activity level. Higher level activity of the road was assumed to be affecting vehicle users awareness of being monitored by the law enforcement. In this analysis, a product moment correlation coefficient (or simply known as correlation coefficient) was calculated using equation (2)

$$r = \frac{1}{n} \sum_{i=1}^n \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{S_x S_y} \tag{2}$$

where

- \bar{X} is the mean value of X
- \bar{Y} is the mean value of Y
- S_x is the standard deviation of X
- S_y is the standard deviation of Y

It was also important to evaluate the significance level of the calculated correlation coefficients. In this study a significance level of 0.05 was used. There were 4 relationships analyzed. i.e. between the safety belt use rate of:

- all and each vehicle type with the road type code.
- all and each vehicle type with the road width.
- all and each vehicle type with V/C at the busiest direction of the divided roads
- all and each vehicle type with V/C at busiest direction of the undivided roads

3. Results

3.1. Safety Belt Use Rate and V/C

In order to provide general description of safety belt use rate across different road types and different vehicle types, Tables 5 and 6 provide such information for drivers and front seat passengers respectively.

Table 5. Safety Belt User Rate of the Drivers

Road Link	Road Type	Safety Belt Use Rate (%)							
		A	B	C	D	E	F	G	All
Otista	6/2 D	96	79	1	93	0	22	91	83
TMII	6/2 D	76	92	0	84	14	23	82	74
Halim PK	4/2 D	87	86	0	91	-	-	84	87
Raya Bogor	4/2 D	87	79	0	86	0	22	68	58
Cawang Baru	2/2 UD	81	82	0	65	-	-	83	53
Condet	2/2 UD	97	88	-	64	-	-	88	85
Pinang Ranti	2/2 UD	80	81	0	79	-	-	73	70

Table 6. Safety Belt User Rate of the Front Seat Passengers

Road Link	Road Type	Safety Belt Use Rate (%)							
		A	B	C	D	E	F	G	All
Otista	6/2 D	73	41	0	64	-	-	79	57
TMII	6/2 D	57	40	0	54	-	-	54	54
Halim PK	4/2 D	75	83	0	63	-	-	69	72
Raya Bogor	4/2 D	58	16	0	43	-	-	40	47
Cawang Baru	2/2 UD	54	29	0	42	-	-	80	23
Condet	2/2 UD	54	17	-	34	-	-	40	48
Pinang Ranti	2/2 UD	56	63	0	65	-	-	47	47

- A. Passenger car (private)
- B. Taxi
- C. Paratransit
- D. Pick-up/Box
- E. Public Bus
- F. Non-public bus
- G. Truck

It should be noted that number of sample was varied between different vehicle types and road links. Therefore the weighted average of safety belt use rate presented in the last columns of Tables 5 and 6 reflect these differences. For example although in all cases the number of sample of the passenger car was the highest compare to the number sample of the other type of vehicle (including the paratransit), the numbers of sample of the paratransit at Raya Bogor and Cawang Baru were significantly high and therefore their zero safety belt use rates decrease the all-vehicle safety belt use rates significantly (Table 5). On the contrary, at TMII the number of sample of the passenger car was significantly higher than number of sample of the paratransit and therefore the paratransit zero safety belt use rate does not decrease the all-vehicle safety belt use rate significantly (Table 5). One should also aware that number of drivers was equal to number of vehicles, but number of front seat passengers was less than number of vehicles. In Tables 5 and 6, “0” means that none of the observed objects were wearing seat belt, whilst “-” means that the observed objects were not available.

The V/C of each road link is presented in Table 7. At divided road, A-B means the busiest direction where safety belt use rate observations were made, whilst B-A was the opposite direction.

Table 7. The V/C of the Observed Links

Road Link	Road Type*	V/C	
		A-B	B-A
Otista	6/2 D	0.77	0.39
TMII	6/2 D	0.37	0.28
Halim PK	4/2 D	0.63	0.36
Raya Bogor	4/2 D	0.95	0.70
Cawang Baru	2/2 UD	0.29	
Condet	2/2 UD	1.00	
Pinang Ranti	2/2 UD	0.45	

- * 6/2 D : 6 lanes-2 ways divided
- 4/2 D : 4 lanes-2 ways divided
- 2/2 UD : 2 lanes-2ways undivided

3.2. Correlation Analysis

Table 8 shows results of the correlation analysis between safety belt use rate of all and each vehicle type with the road type code.

Table 8. Correlation between the Safety Belt Use Rate and the Road Type Code

Vehicle Type	Driver		Passenger	
	r	Significant at $\alpha=0.05$?	r	Significant at $\alpha=0.05$?
All	0.31	no	0.43	no
Passenger car	0.01	no	0.55	no
Taxi	0.14	no	0.10	no
Paratransit	0.55	no	0.00	no
Pick-up/Box	0.52	no	0.43	no
Public Bus	0.50	no	-	-
Non-Public Bus	0.28	no	-	-
Truck	0.78	yes	0.26	no

Table 9 shows results of the correlation analysis between safety belt use rate of all and each vehicle type with the road width.

Table 9. Correlation between the Safety Belt Use Rate and the Road Width

Vehicle Type	Driver		Passenger	
	<i>r</i>	Significant at $\alpha=0.05$?	<i>r</i>	Significant at $\alpha=0.05$?
All	0.23	no	0.33	no
Passenger car	0.04	no	0.52	no
Taxi	0.02	no	0.03	no
Paratransit	0.71	no	0.00	no
Pick-up/Box	0.70	no	0.43	no
Public Bus	0.40	no	-	-
Non-Public Bus	0.23	no	-	-
Truck	0.35	no	0.40	no

Table 10 shows results of the correlation analysis between safety belt use rate of all and each vehicle type with V/C at the busiest direction of the divided roads.

Table 10. Correlation between the Safety Belt Use Rate and the V/C at the Busiest Direction of the Divided Roads

Vehicle Type	Driver		Passenger	
	<i>r</i>	Significant at $\alpha=0.05$?	<i>r</i>	Significant at $\alpha=0.05$?
All	-0.43	no	-0.66	no
Passenger car	0.70	no	-0.80	no
Taxi	-0.95	yes	-0.36	no
Paratransit	0.25	no	0.00	no
Pick-up/Box	0.32	no	-0.91	no
Public Bus	-0.95	yes	-	-
Non-Public Bus	0.11	no	-	-
Truck	-0.48	no	-0.98	yes

Table 11 shows results of the correlation analysis between safety belt use rate of all and each vehicle type with V/C at the busiest direction of the undivided roads.

Table 11. Correlation between the Safety Belt Use Rate and the V/C at the Busiest Direction of the Undivided Roads

Vehicle Type	Driver		Passenger	
	<i>r</i>	Significant at $\alpha=0.05$?	<i>r</i>	Significant at $\alpha=0.05$?
All	0.93	no	0.69	no
Passenger car	0.96	no	-0.31	no
Taxi	0.94	no	-0.54	no
Paratransit	-	-	-	-
Pick-up/Box	-0.36	no	-0.54	no
Public Bus	-	-	-	-
Non-Public Bus	-	-	-	-
Truck	0.51	no	-0.77	no

4. Discussion

It can be seen from Table 8 that in most cases there were no significant correlation between safety belt use rate and road type code. From Table 9, it is clear that there were no significant correlation between safety belt use rate and road width. Therefore road type code and road width were not the valid representations of road activity level that was assumed to be affecting general vehicle users awareness of being monitored by the law enforcement officers. However the only significant correlation between safety belt use rate and road type code in Table 8 was for Truck. It might be caused by relatively higher enforcement activity at higher road class that was experienced by goods vehicles. Similar explanation might be applied for the high correlations between safety belt use rate and road width for pick-up.

Table 10 shows that all significant correlations were negative. This suggests that the higher the V/C, the lower the safety belt use rate. As speed is the opposite of V/C, it seems that at a divided road the safety belt use rate increase as the speed increase (perhaps for safety reason not for possible enforcement reason). Therefore V/C was not a valid representation of a divided road activity level that was assumed to be affecting vehicle users awareness of being monitored by the law enforcement officers.

There were only three pairs of data used to produce each correlation coefficient presented in Table 11 (the degree of freedom was only one). Therefore to be significant at $\alpha=0.05$ the *r* should be at least 0.997. Although none of the correlation coefficients was significant at $\alpha=0.05$, there were three correlation coefficients which were quite high. One of them was for the driver of "all vehicle" (0.93). This suggests that at an undivided road the safety belt use rate increase as the V/C increase. Therefore V/C was a valid representation of undivided road activity level that was assumed to be affecting vehicle users awareness of being monitored by the law enforcement officers.

5. Conclusion

The results show that number of lanes and road widths were not suitable to predict safety belt use rate of the driver and the front seat passenger. For two lanes two ways road, a significant and high positive correlation was found between safety belt use rates and V/Cs. On the contrary for four lanes and six lanes two ways divided road, a significant and negative correlation was found between safety belt use rates and V/Cs. This suggests that in wider road, driver awareness of wearing seat belt might be more affected by the speed (an inverse of V/C) due to possibility to drive in higher speed.

5. References

- [1] Shults, R.A., Elder, R.W., Sleet, D.A., Thompson, R.S., Nichols, J.L. 2004. Primary Enforcement Seat Belt Laws are Effective Even in the Face of Rising Belt Use Rates. *Accident Analysis & Prevention* 36: 491-493.
- [2] Ulmer, R.G., Preusser, C.W., Preusser, D.F., Cosgrove, L.A. 1995. Evaluation of California's Safety Belt Law Change from Secondary to Primary Enforcement. *Journal of Safety Research* 26(4): 213-220.
- [3] Escobedo, L.G., Chorba, T.L., Remington, P.L., Anda, R.F., Sanderson, L., Zaidi, A.A. 1992. The Influence of Safety Belt Laws on Self-Reported Safety Belt Use in the United States. *Accident Analysis & Prevention* 24(6): 643-653.
- [4] Streff, F.M., Molnar, L.J. 1991. Use of Automatic Safety Belts in Michigan. *Journal of Safety Research* 22: 141-146.
- [5] Johnston, J.J., Hendricks, S.A., Fike, J.M. 1994. Effectiveness of Behavioral Safety Belt Interventions. *Accident Analysis & Prevention* 26(3): 315-323.
- [6] Shinar, D. 1993. Demographic and Socioeconomic Correlates of Safety Belt Use. *Accident Analysis & Prevention* 25(6): 745-755.
- [7] Calisir, F., Lehto, M.R. 2002. Young Driver's Decision Making and Safety Belt Use. *Accident Analysis & Prevention* 34: 793-805.
- [8] Hong, S., Kim, D., Kritkuasky, K., Rashid, R. 1998. Effects of Imitative Behaviour on Seat Belt Usage: Three Field Observational Studies. *Proceedings of the Human Factors Society* 2: 1093-1097.
- [9] Shin, D., Hong, L., Waldron, I. 1999. Possible Causes of Socioeconomic and ethnic Differences in Seat Belt Use Among High School Students. *Accident Analysis & Prevention* 31(5): 485-496.
- [10] Putranto, L.S. 2005. Characteristics of Seat Belt Use in Jakarta. *Proceeding of the 6th International Conference of the Eastern Asia Society for Transportation Studies*.

Laboratory Evaluation of Horizontal Coefficient of Consolidation of Fibrous Peat Soil

Yulindasari Sutejo¹, Wong Leong Sing², Nurly Gofar³

Graduate Student¹, Former Graduate Student² and Senior Lecturer³

Department of Geotechnics and Transportation, Faculty of Civil Engineering,
Universiti Teknologi Malaysia, 81310 Skudai, Johor
E-mail: indatejo@yahoo.com

Abstract

Encountered extensively in wetlands, fibrous peat is considered as problematic soil because it exhibits unusual compression behavior. When a mass of fibrous peat soil with both vertical and horizontal drainage boundaries is subjected to a consolidation pressure, rate of excess pore water dissipation from the soil in the horizontal direction (c_h) is expected to be higher than that in the vertical direction (c_v). The ratio of c_h/c_v of two is commonly used in practice for estimation of consolidation in soft clay improved by vertical drain. Published data showed that the c_h/c_v ratio for fibrous peat could be as high as 300. This research focused on the consolidation behavior of fibrous peat from Kampung Bahru, Pontian, Johor with respect to one-dimensional vertical and horizontal consolidation. The result is useful for evaluation of the utilization of some type of vertical drainage for soil improvement to accelerate the settlement of fibrous peat soil. Results from constant head permeability reveal that the soil is almost isotropic as indicated by equal initial permeability in horizontal and vertical direction. Hydraulic consolidation tests in Rowe cell, the coefficient of rate of horizontal consolidation increases significantly with consolidation pressure, while the increase in the coefficient of rate of vertical consolidation does not increase as much. The finding shows that the c_h/c_v ratio increases as consolidation pressure increases. The c_h/c_v ratio increase from 3.5 to 6 for consolidation pressure of 25 to 200 kPa. The ratio of coefficient of permeability k_h/k_v under a consolidation pressure of 200 kPa is about 5. This finding implies that the utilization of horizontal drain maybe suitable for accelerating settlement and reducing the effect of secondary consolidation.

Keywords:

Fibrous Peat, Horizontal Consolidation, Permeability

1. Introduction

Peat is usually found as an extremely loose, wet, unconsolidated surface deposit which forms as an integral part of a wetland system, therefore; access to the peat deposit is usually very difficult as the water table exists at, near or above the ground surface. This type of soil covers large area in West Johore including Pontian, Batu Pahat and Muar. As part of the development in this area, many civil engineering structures have to be constructed over peat deposit. There are many methods available for construction and stabilization of the peat such as: soil replacement, surface reinforcement, preloading, chemical stabilization, sand or stone column, pre-fabricated vertical drains, and pile (e.g. Huat, 2004). The selection of the most appropriate method should be based on the examination of the characteristics of the soil at a specific location because research has established that the behavior of peat soil is site specific.

Fibrous peat is a mixture of fragmented organic material formed in wetlands under appropriate climatic and topographic conditions. The soil has essentially an open

structure with interstices filled with a secondary structural arrangement of nonwoody, fine fibrous material (Dhowian and Edil, 1980). According to Karlsson and Hansbo (1981), fibrous peat differs from amorphous peat in that it has a low degree of decomposition, fibrous structure, and easily recognizable plant structure. The compressibility of fibrous peat is very high and so the rate of consolidation.

The formation of peat deposit led to a pronounced structural anisotropy in which the fibers tend to have horizontal orientation. The particle arrangement influences the rate of flow as water tries to dissipate from soil under loading. Thus, even though the application of load tends to reduce the rate of consolidation, the effect is not as pronounced for horizontal flow of water. In other words, under consolidation pressure, water tends to dissipate faster from the soil if they are allowed to flow in horizontal direction.

General practice is to use coefficient of rate of horizontal consolidation (c_h) twice the coefficient of rate of vertical consolidation (c_v) for clay and the ratio is much higher for peat soil. Parallel to the coefficient of rate of consolidation, Dhowian and Edil (1980) suggested that for predominantly fibrous peat soils, the horizontal hydraulic conductivity (k_h) is greater than that in vertical direction (k_v)

* Corresponding Author. E-mail: indatejo@yahoo.com,
Tel: +60-7-5532517, Fax: +60-7-5566157

by an order of magnitude. The subsequent research by Edil et al. (2001) has shown that for peat with high fiber content, the ratio of c_h/c_v could be as high as 300.

The research focuses on the study of compressibility of fibrous peat due to primary consolidation or dissipation of excess pore water pressure in horizontal direction, and the effect of secondary consolidation on the horizontal coefficient of consolidation, c_h . The effect of consolidation pressure on the vertical and horizontal coefficient of permeability, (k_h and k_v) are also the subject of the study. The results are expected to give practical information for the selection of soil improvement method for construction on fibrous peat soil to be used in the study area.

2. Experimental Setup

The study was carried out on fibrous peat obtained from Kampung Bahru, Pontian, Johor. The samples were acquired using blocks sampling method whereby the soil was excavated to a depth of about 1 m and then a tube of 300 mm-diameter and 300 mm high was pushed slowly into the soil. The surrounding of the tube was excavated to enable the cutting of the bottom of the sample. The quality of samples was maintained by ensuring the sharpness of the edge of the tube and knife used to cut the sample and by restoring the soil and water using wax and wooden plate cover before they were transported to the laboratory. Eighteen samples were obtained from six different points, at least 2 meter apart in one location. Each sample was transported in a well cushioned wooden box and was kept in the laboratory under constant temperature (air conditioned room).

Index soil properties and classification tests were conducted in order to determine the physical characteristics and to classify the peat. Even though most of the features of anisotropy of the fibrous peat are visible to the naked eye, a more detailed analysis on the structural arrangement and microstructure of the fiber and the fiber content can be examined under a Scanning Electron Microscope (SEM). Fiber content and fiber arrangement is known to have an effect on the compression behavior of peat especially those related to permeability or the ability of water to flow through the soil. Evaluation of permeability is made in this research for evaluation of the effect of fiber on the initial permeability and effect of application of consolidation pressure on the reorientation of the fiber and permeability of the soil. Constant-head permeability tests were carried out to determine initial hydraulic conductivity of the peat, while data from oedometer consolidation tests were used to determine the pre-consolidation pressure (σ_p), to decide the range of consolidation pressures for hydraulic consolidation tests, and to observe the long-term compression behavior of the peat.

The focus of the study was to evaluate the coefficient of rate of consolidation of fibrous peat based on the results of consolidation tests conducted on Rowe cell under a range of consolidation pressures (Figure 1). The cell has an internal diameter of 150 mm and height of 50 mm, and the soil samples were subjected to hydraulic consolidation pressures

of 25, 50, 100, and 200 kPa. Two types of hydraulic consolidation tests were performed in Rowe cell, namely consolidation test with two-way vertical drainage and test with radial drainage to periphery.



Figure 1. Rowe Consolidation Cell

Designation of the hydraulic consolidation test with two-way vertical drainage is shown in Figure 2. In this type of test, drainage takes place at both top and bottom faces of the sample. Porous drainage discs are placed at the top and bottom of the sample; both are connected to the same back pressure system for consolidation stages. The top disk also functions as a means to provide equal strain and top drainage from the sample.

Designation of the hydraulic consolidation test with radial drainage to periphery is shown in Figure 3. In this case, the sample is surrounded by a drainage layer of porous plastic material. The top surface of the sample is covered by an impermeable steel disc. A back pressure system with volume gauge is connected to the rim drain at the top of the cell, via valve F. Pore water pressure is measured at the base of the sample from the centre. The pore pressure transducer housing block is connected to valve A which replaces the blanking plug at that cell outlet.

Finally the effect of consolidation pressure on fabric arrangement and therefore the permeability of the peat were studied by carrying out permeability test on Rowe cell under consolidation pressure of 100 and 200 kPa for vertical and horizontal drainage.

All laboratory tests are carried out following the procedures outlined in the manual of soil laboratory testing (Head, 1981, 1982, 1986) in accordance with the British Standards (BS) and American Standard Testing Methods (ASTM).

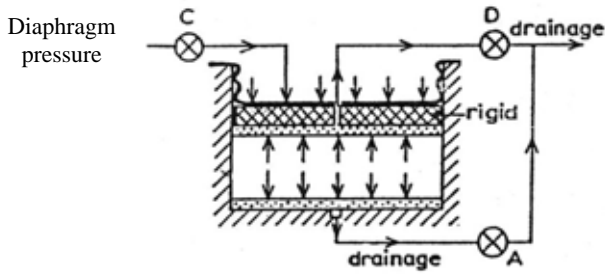


Figure 2. Two-Way Vertical Drainage and Loading Condition for Hydraulic Consolidation Test in Rowe Cell with 'Equal Strain' Loading (Head, 1986)

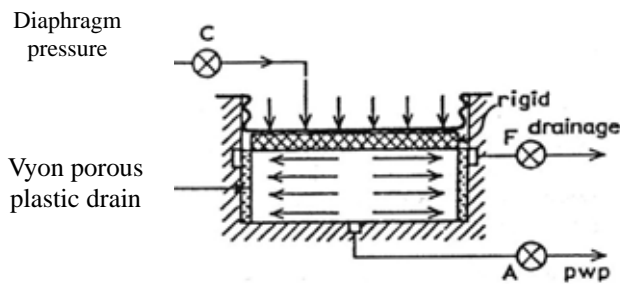


Figure 3. Radial Drainage to Periphery, and Loading Condition for Hydraulic Consolidation Test in Rowe Cell with 'Equal Strain' Loading (Head, 1986)

3. Data Analysis

Graphical plots of settlement, volume change, and pore-water pressure as a function of time were obtained from each loading stage of a Rowe cell consolidation test. These graphs are used to determine the time corresponding to primary consolidation and secondary compression, from which the coefficient of consolidation can be calculated by using an equation with an appropriate multiplying factor.

Robinson's (2003) method is used for the analysis of the time-compression curve because it is suitable for the analysis of time-compression curves in which the secondary consolidation starts during the dissipation of pore-water-pressure. The secondary compression index and the beginning of secondary compression are among the parameters required for the analysis of secondary compression. The pore pressure dissipation graph which gives a better indication of t_{50} and the corresponding T_{50} is to be used for evaluation of coefficient of rate of consolidation (c_h and c_v).

Besides the time compression curve, a graph relating the void ratio at the end of each loading stage with the effective pressure on a linear or logarithmic scale was plotted for a complete set of consolidation test data. The $e-p'$ curve is used to obtain the coefficient of axial compressibility a_v and thus the coefficient of volume compressibility m_v , while the $e-\log p'$ is used to obtain compression and recompression indexes, c_c and c_r respectively. These data are required for evaluation of the magnitude of primary settlement and to

obtain the ratio of c_d/c_c for calculation of secondary compression of the peat.

Permeability in horizontal and vertical directions (k_h and k_v) are calculated based on the coefficient of consolidation or tested on Rowe cell upon the completion of the consolidation test. The values were evaluated and compared with the initial coefficient of permeability obtained from constant head permeability test.

Comparison between c_h and c_v as well as comparison between k_h and k_v under various consolidation pressure was evaluated to confirm the hypothesis developed for the study that the dissipation of pore water pressure in horizontal direction is actually faster than that in vertical direction. The ratios of c_h/c_v and k_h/k_v can be used for analysis on the effect of fabric on the properties.

4. Results and Discussion

4.1 Preliminary data

Data on the fundamental properties of the peat soil is summarized in Table 1. As shown in Table 1, the peat soil is acidic and has high organic and fiber contents. Average moisture content of 608% indicates that the peat has a high water-holding capacity. Based on von Post humification scale, the peat can be classified as H_4 or low to medium degree of decomposition. The data allowed the peat under study is a typical of fibrous peat found in West coast of Peninsular Malaysia (Muttalib, 1991 and Huat, 2004).

Table 1. Basic Properties of the Peat Soil

	Parameters	Results	Published data (ranges)
Index properties	Von Post humification of peat	H_4	$H_1 - H_4$
	Natural water content (%)	608	200 – 700
	Bulk unit weight (kN/m^3)	10.02	8.30 – 11.50
	Dry unit weight (kN/m^3)	1.40	
	Specific Gravity (G_s)	1.47	1.30 – 1.80
	Initial void ratio (e_o)	8.92	3 – 15
	Acidity (pH)	3.24	3.0 – 4.5
Classification	% < 0.063 mm	2.74	
	Organic content (%)	97	> 90
	Ash content (%)	3	< 10
	Fiber content (%)	90	> 20

Figure 4 shows the typical fiber orientation obtained by Scanning Electron Microscope for the fibrous peat at initial state and under consolidation pressure of 200 kPa. The samples were cut in vertical and horizontal sections to enable the observation of the rearrangement of the fiber due to application of consolidation pressure. From the results of SEM (Figure 4), it can be observed that the initial arrangement of fiber is somewhat uniform and the effect or roots is visible. Thus, at initial stage the water may take an erratic drainage path resulting in almost uniform permeability. Comparison of the four pictures indicates a pronounced structural anisotropy for the fibrous peat with the void spaces in the horizontal direction larger than those in the vertical direction resulting from the fiber orientation within the soil. Individual microstructures remained essentially intact after compression under high-stress conditions.

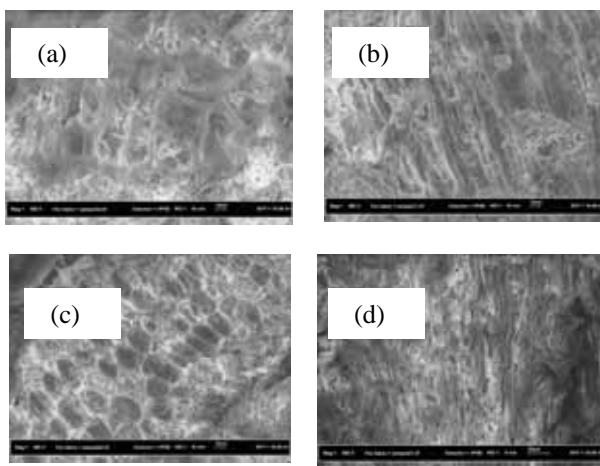


Figure 4. SEM of Fibrous Peat Samples (A) Horizontal Section at Initial State X 400, (B) Vertical Section at Initial State X 400, (C) Horizontal Section under Consolidation Pressure of 200 kPa X 400, (D) Vertical Section under Consolidation Pressure of 200 kPa X 400

Standard Oedometer test was conducted on 12 samples. Each of the samples has a thickness of 20.13 mm, a diameter of 50.23 mm, and was subjected to consolidation pressures of 12.5 kPa, 25 kPa, 50 kPa, 100 kPa, 200 kPa, and 400 kPa. The results indicate that primary consolidation and secondary compression characteristics of the soil can be identified from consolidation curves. The primary consolidation is dominant in the compression of the peat due to high initial void ratio, but the consolidation occurs in a relatively short time as compared to clay. Secondary compression, even though less significant than the primary consolidation in term of magnitude, could be very important in term of the design life of a structure. Tertiary compression was observed from the test results, but may not be very significant in term of the design life of a structure because the secondary compression takes a significant amount of time.

The e-log p' curve obtained from the set of data shows that the pre-consolidation pressure is about 45 kPa and the compression index is 3.772. Based on the Oedometer data,

the range of consolidation pressure to be used in large strain consolidation test is 25, 50, 100, and 200 kPa.

4.2. Consolidation Characteristics

Five sets of vertical and radial consolidation tests were conducted on 'identical' samples. The sample was placed on a Rowe consolidation cell with diameter of 151.4 mm and height of 48.78 mm and subjected to hydraulic consolidation pressures of 25, 50, 100, and 200 kPa during the test. Figure 5 shows the typical time-compression curve obtained from the results of consolidation test. The shape of log time-compression curve indicates that deformation process of fibrous peat did not give a clear indication of an inflection point where the primary consolidation is assumed to end and the secondary consolidation is assumed to start. The secondary consolidation may have started during the process of pore water pressure dissipation. Therefore, Robinson's (2003) method is used for the analysis of time-compression curve.

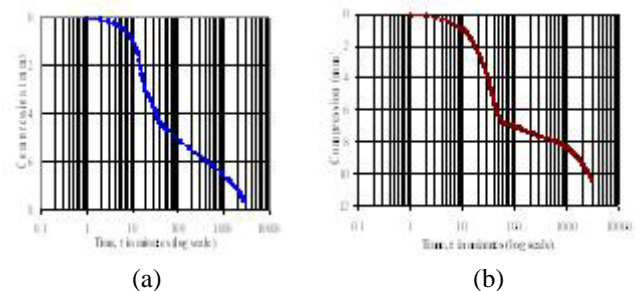


Figure 5. Typical Time-Compression Curve for (a) Vertical; (b) Horizontal Consolidation

Comparison of time-compression curves obtained from consolidation test with vertical and horizontal drainage shows that the secondary consolidation started earlier when subjected to vertical drainage as compared to horizontal drainage. This might be due to the fact that the fiber tends to rearrange in horizontal direction during primary consolidation that the end of the primary consolidation with horizontal drainage is more visible when compared to the consolidation with two-way vertical drainage.

The secondary compression seems to be more dominant when the soil is subjected to consolidation pressure with vertical drainage. Analysis of test results indicated that the secondary compression started after 65% and 80 % degree of consolidation when subjected to vertical and horizontal drainage respectively. The secondary compression started later when the water is allowed to flow in horizontal direction. This means that the effect of secondary compression on primary consolidation can be minimized by providing vertical drain in fibrous peat deposit.

There is no clear trend on the relationship between the coefficients of secondary consolidation with consolidation pressure. The coefficient of secondary consolidation in vertical direction is 0.261 while the secondary consolidation in horizontal direction is 0.226. The e-log p' curve obtained from the test results showed that the compression index c_c of

the soil is 3.128 and 2.879 for vertical and horizontal consolidation respectively. The c_d/c_c ratio calculated from the above data is 0.08 which is slightly higher than published data on different type of peat (Ajlouni, 2000). This might be due to the high fiber content of the soil.

Based on Robinson's (2003) method, the coefficient of consolidation should be calculated based time compression curve for primary consolidation only. The results for c_v and c_h and their variation with consolidation pressure are shown in Figure 6. It is clear from the figure that the coefficient of consolidation decreases almost linearly with increasing consolidation pressure, however the effect is more significant for samples subjected to vertical consolidation, due to the rearrangement of the fiber. The ratio of c_h/c_v obtained based on the average values are shown in Figure 7. A ratio of about 3.5 was obtained even under a low consolidation pressure of 25 kPa and the ratio increases as the consolidation pressure increases. A ratio of 5.89 was obtained for consolidation pressure of 200 kPa.

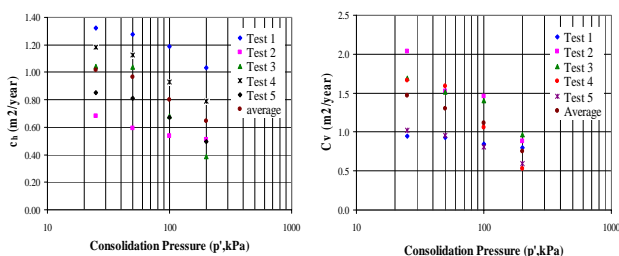


Figure 6. Variation of Coefficient of Rate of Consolidation with Consolidation Pressure for Sample Tested under (a) Vertical Consolidation (b) Horizontal Consolidation

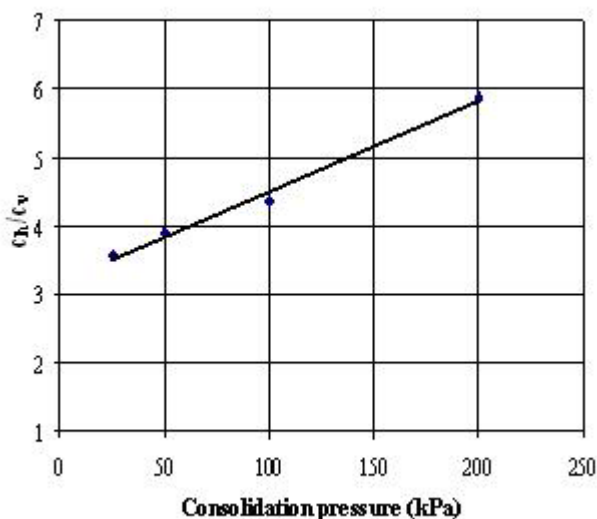


Figure 7. Ratio of Coefficient of Consolidation for a Range of Consolidation Pressure

4.3 Permeability

The rate of consolidation of the fully saturated and undisturbed fibrous peat soil is controlled by the permeability of the soil. Four sets of constant head permeability test were conducted on samples obtained by pushing a piston sampler into the soil in vertical and horizontal directions.

The results of the constant head permeability test revealed that at initial state, the average vertical and horizontal coefficient of permeability of the soil at standard temperature of 20°C is 1.20×10^{-4} m/s and 9.48×10^{-5} m/s respectively. Initial permeability of the peat is very high and it is comparable to sand. The results also suggested that at initial state, the average horizontal coefficient of permeability is actually slightly lower than the average vertical one, and the ratio of k_h/k_v at initial state is 0.79.

It has been established that the permeability is highly affected by compression or reduction in void ratio. To study the effect of consolidation pressure on the coefficient of permeability, permeability test was carried out at a consolidation pressure of 200 kPa. The result shows a significant decrease in the permeability in which the coefficients of permeability under 200 kPa in vertical and horizontal directions are 5.07×10^{-10} m/s and 26×10^{-10} m/s respectively. The ratio of k_h/k_v under consolidation pressure of 200 kPa is 5.13.

Coefficient of permeability was also evaluated based on the coefficient of consolidation obtained from consolidation test for consolidation pressure of 50, 100, and 200 kPa data show a decrease in permeability as the consolidation pressure is increases. Combination of the data from various tests indicates an increase in the ratio of k_h/k_v as shown in Figure 8.

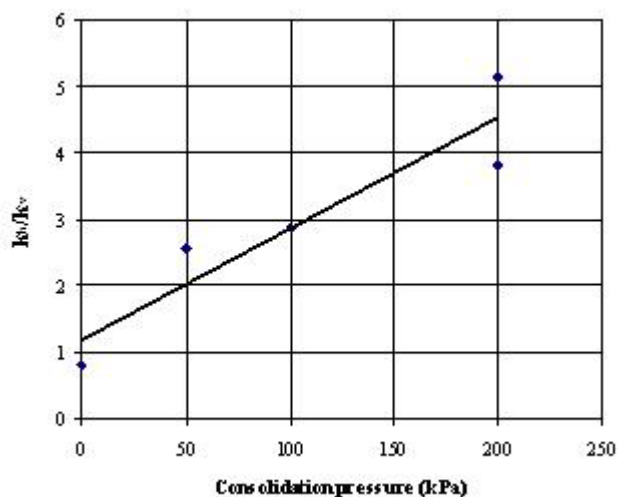


Figure 8. Permeability Ratio (k_h/k_v) for a Range of Consolidation Pressure

4. Conclusions

Several conclusions can be derived from the study on the horizontal coefficient of consolidation of fibrous peat found in Kampung Bahru, Pontian, West Johore:

1. The fibrous peat has a high compressibility with significant secondary compression stage. The soil has a relatively short duration of primary consolidation.
2. Observation on the time-compression shows that the secondary compression started later when the water is allowed to flow in horizontal direction. Lower coefficient of consolidation was obtained from the test with horizontal drainage, but the ratio of c_h/c_c is almost equal (0.08).
3. The coefficient of rate of consolidation is higher in the horizontal direction than in the vertical direction.
4. Application of consolidation pressure has the effect of decreasing the coefficient of consolidation; however the horizontal rate of consolidation decreases in a slower rate than the vertical coefficient of consolidation due to the rearrangement of the fiber in the soil. The ratio of c_h/c_v is increasing from 3.5 to 5.89 for consolidation pressure of 25 to 200 kPa.
5. Consolidation pressure causes a rearrangement of fiber in the vertical direction more than in horizontal direction. The k_h/k_v ratio increases from 0.79 for initial condition to about 5 under consolidation pressure of 200 kPa.
6. Utilization of vertical drainage has a positive effect on speeding up the primary consolidation process, and minimizing the effect of secondary compression during primary consolidation, thus; the post-construction estimation of the post-construction settlement could be made based only on the secondary compression.

5. Acknowledgement

The study is part of UTM Fundamental research under vot 75210 entitled "Determination of Coefficient of Rate of Horizontal Consolidation of Peat Soil"

6. References

- [1] Ajlouni, M.A. 2000. *Geotechnical Properties of Peat and Related Engineering Problems*. PhD Thesis. University of Illinois at Urbana-Champaign.
- [2] Dhowian, A.W., and Edil, T.B. 1980. Consolidation Behavior of Peats. *Geotechnical Testing Journal*. 3(3): 105-114.
- [3] Edil, T.B. 2001. Site Characterization in Peat and Organic Soils. In *Proceeding of the International Conference on In Situ Measurement of Soil Properties and Case Histories*, 49-59, Bali, Indonesia.
- [4] Hartlen, J., and Wolski, W. (Eds.) 1996. *Embankments on Organic Soils*. Amsterdam: Elsevier Science B.V.
- [5] Head, K.H. 1980, 1982, and 1986. *Manual of Soil Laboratory Testing*, Vol. 1, 2, and 3, Pentech Press Ltd, London.
- [6] Huat, Bujang, B.K., 2004. *Organic and Peat Soil Engineering*. Univ. Putra Malaysia Press.
- [7] Robinson, R. G. 2003. A Study on the Beginning of Secondary Compression of Soils. *Journal of Testing and Evaluation*. 2003. 31(5): 1-10.
- [8] Rowe, P.W. and Barden, L. 1966. New Consolidation Cell. *Geotechnique*. 16: 162-169.

Influence of Natural Organic Matter Characteristics on Fouling and Rejection Mechanisms of Immersed Ultrafiltration Membrane filtration. A Case Study of Potable Water Production

A.W. Zularisam¹, A.F. Ismail^{2*}, Razman Salim³, M. Sakinah^{1,2}

¹Environmental Engineering Focus Group, Faculty of Civil Engineering & Environmental Engineering, Kolej Universiti Kejuruteraan & Teknologi Malaysia, KUKTEM, Gambang, Pahang, Malaysia

²Membrane Research Unit, Universiti Teknologi Malaysia, Skudai Johor, Malaysia

³Faculty of Civil Engineering, Universiti Teknologi Malaysia, Skudai Johor, Malaysia

Abstract

Natural organic matter (NOM) has been claimed as the main precursor for disinfection by product (DBPs) that is carcinogen to human health. Chlorine which is used as disinfectant in drinking water post treatment will react with NOM to produce chlorinated organic such as trihalomethane (THM), haloacetic acid (HAAs) and other halogenated by products. Besides that, NOM is also found to act as the major membrane foulant in drinking water treatment processes. This phenomenon has caused inevitable fouling problems which resulted in low efficiency of membrane performance. However, the primary fraction of NOMs, which is believed to cause membrane fouling is still unclear. The main objective of this study is to evaluate the effect of fractionated NOM and their fouling potentials on the submerged ultrafiltration membrane reactor performance. The membrane fouling characteristics of NOM based on hydraulic resistances (R_{cp} , R_g , R_{wa} , R_{sa}), rejection of UV_{254nm} , DOC, and flux decline were assessed using 68 kDa MWCO hydrophobic polysulfone (PS) and 50 kDa MWCO hydrophilic cellulose acetate (CA) membranes. Both PS and CA membranes were spun by a simple forced convective dry/wet spinning technique. The Sungai Pontian Kecil river water was taken as a sample and was subjected to non-ionic macro-porous ion exchange resins (DAX-8 and XAD-4). The sample was fractionated into three specific classes which were hydrophilic, transphilic and hydrophilic compounds. Results showed that the low aromatic hydrophilic fraction content of NOM attributed to the worst flux decline despite lesser solute rejection than hydrophobic component. In addition, the PS membrane was surprisingly found to exhibit higher NOM rejection especially for the component with higher SUVA (UV_{254nm}/DOC) despite possessed higher MWCO than the hydrophilic CA membrane. Charge interaction between hydrophobic component and membrane surface is hypothesized in playing important role than steric hindrance mechanism. In fact NOM rejection mechanisms were found to be a function of both membrane type and water characteristics.

Keywords: NOM fraction; fouling; submerged hollow fiber ultra filtration

1. Introduction

Membrane filtration process does offer extra advantages over conventional treatment such as small footprint, compact module and capable of handling wide fluctuations in feed quality. MF and UF are employed to remove microparticles and macromolecules, which generally include inorganic particles, organic colloidal (i.e. microorganism) and dissolved organic matters (DOM). DOM is ubiquitous in natural surface water and often reclaimed [1-4] as important factor for both the reversible and irreversible fouling in water filtration. Unfortunately the presence of free chlorine content that is used as a disinfectant in the conventional treatment is found to react with this residual DOM consisting in the natural organic matter (NOM). This reaction process has been found to have tendency to form disinfection by product (DBPs). In this study the influence of NOM components characteristics to the fouling of a submerged ultrafiltration polysulfone membrane and fouling mechanisms were investigated. The potential foulants of NOM isolates which were detrimental

to the employed submerged UF membrane were studied and identified by isolating the chosen surface water. The performance of membrane and the associated fouling mechanisms were examined based on NOM isolates rejection, flux declination, hydraulic series resistance.

2.0 Approach and Methods

2.1 Membrane Fabrication

The formulated dope was then spun by using a forced convective dry/wet spinning technique under pressurized nitrogen gas. The membrane characteristic is shown in Table 1. The polysulfone polymer was supplied by Amoco Performance Product Inc. The additive polymer of PVP3000 was bought from Fluka Milwaukee while the solvent (DMEC) was purchased from Merck Darmstadt Germany. The PVP3000, solvent and PSF were sequentially mixed in a container until a homogenous mixture was achieved at a constant temperature of 60 °C.

*Corresponding author email: afauzi@utm.my

Table 1. Characteristics of the experimental membranes

Parameter	MRUTM55	
MRUTM66		
Type	Hollow fiber	Hollow fiber
Material	Polysulfone (PSF)	Cellulose
Acetate		
Contact angle	56°	20°
MWCO	68 kDA	50 kDA

Table 2. Surface water characteristics

Source	Sungai Ulu Pontian, Johor	
DOC (mg/L)	6.8	
UV ₂₅₄ (cm ⁻¹)	0.088	
SUVA (L/mg.m)	1.29	

Table 3. DOC concentration of Sungai Ulu Pontian fractions (based on DOC and mass balance technique)

Fraction (%)		DOC (mg/L)	DOC
Hydrophobic (HPO)		2.38	35
Transphilic (TPI)		1.7	25
Hydrophilic (HPI)		2.6	38

2.2 Water Quality

A soft and rich in NOM surface water (Table 2) was taken from Sungai Ulu Pontian Johor, Malaysia. The raw water was filtered by a 0.45 µm and fractionated into hydrophobic (HPO) fraction which is Supelite DAX-8 (Sigma-Aldrich) adsorbable, transphilic (TPI) which is Amberlite XAD-4 adsorbable and hydrophilic (HPI) components which pass through the DAX-8 and XAD-4 resins without any adsorption (Table 3). The HPO and TPI components were collected after eluting the resins with 0.1 NaOH.

2.3 Determination of Hydraulic Resistance

Ultrafiltration (UF) submerged reactor was used throughout this study and the experimental set-up is

schematically illustrated in Figure 2. Initial flux of pure water permeate (J_{pwp}) was introduced to the membrane at TMP of 140 mmHg until the pure water permeability was stabilized. The DI water was replaced with a NOM fractional source at the same TMP of J_{pwp1} . The operational flux and flux decline were regularly measured throughout the experiment. Samplings for the UV₂₅₄ and DOC rejection were periodically taken until permeation reaches steady state conditions. The system was then replaced with DI water to remove (CP) layer and the second J_{pwp2} was measured. The system was back flushed thoroughly for 10 minutes [2] with DI water to remove all reversible adsorption layers before the third J_{pwp3} was measured and the membrane was taken out from the reactor. The foulants were extracted using 0.1 N NaOH for 3 hours followed by 0.1N HCl for another 3 hours solutions respectively [3]. The fourth J_{pwp4} was then measured to complete the Resistance in series model (Equation 1). The assumption on this model is made whereby the osmotic pressure is subsumed into the concentration polarization resistance.

$$J = \frac{\Delta P}{\mu(R_m + R_{cp} + R_c + R_a)} \quad (1)$$

where R_m is membrane intrinsic hydraulic resistance, R_{cp} is concentration polarization resistance, R_c is cake resistance, R_a is adsorption resistance.

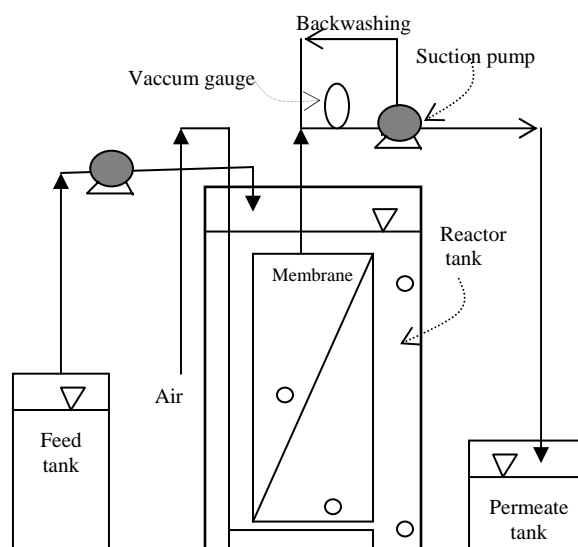


Figure 2. Schematic diagram of submerged UF reactor

3.0 Results and Discussion

3.1 Influence of NOM Fractions On Fouling Characteristics

The Sungai Ulu Pontian exhibited a relatively low hydrophobic fraction concentration and therefore can be classified as a hydrophilic surface water type (Table 3). Surface water with SUVA less than 3 Lm⁻¹mg⁻¹ is classified as low-humic water [4]. The

normalized permeate fluxes for the three fractions are shown in Figure 3. All the three fractions showed an obvious flux decline at the initial period of filtration. The fouling occurred in these cases were mostly attributed by the dissolved NOM.

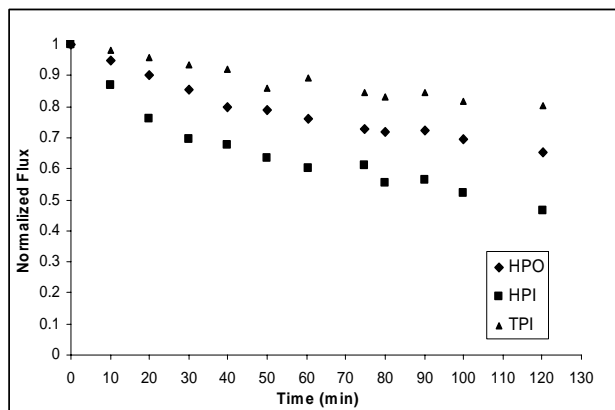


Figure 3. Flux profile of Sungai Ulu Pontian fractions by MRUTM55

The HPI fraction was found to perform the most significance fouling potential than other fractions and this finding could be taken as the most outstanding phenomenon observed in this study. The detrimental effect of HPI fraction to membrane permeability was even more pronounced as it showed greater flux decline and higher fouling rate than HPO and TPI components. However, this finding was contradictory with the results obtained by [5] where in their studies, the HPI components showed less fouling effect than humic fraction. Nevertheless the result presented here showed that the HPI exhibited the worst flux decline by approximately 54% of the initial flux after 120 minutes operation. On the other hand both HPO and TPI flux decline were about 35% and 20% in spite of having the same filtration duration. Thus it is worth noting that the HPI fraction was the prime foulant in NOM and was mainly responsible for fouling and major flux decline in surface water filtration. This could be reasoned from the low UV absorbing compound and high molecular weight component such as polysaccharides or polysaccharides types contained by the HPI fraction. This finding was well supported by studies carried out by done other researcher [6] where they also found that the hydrophilic components cause more fouling potential than humic substances. Although the flux results showed the HPI fraction is the most detrimental foulant and exerted the worst flux decline but yet its DOC removal was the lowest (23%) in among other fractions. In this scenario the rule of thumb whereby normally highest flux decline would performance the best rejection did not occur. In fact it seems that the flux declination does not linearly parallel with the DOC removal and this case was even worsening when the UV_{254} removal was also the least. In this case apparently the HPI fraction is not only the most significant fouling potential but also would result in the worst membrane fouling (flux decline) and the lowest NOM removal (DOC and UV_{254} rejections). In addition the HPO fraction was found to be

well rejected than other fraction in term of DOC and UV_{254} (Figure 4). This phenomenon is presumed to occur due to the electrostatic interaction between HPO fraction and the PSF membrane. HPO fraction contains mainly humic acid and fulvic acid that are anionic polyelectrolyte with negatively charged of carboxylic (COOH), methoxyl carbonyls (C=O) and phenolic (OH). Besides that, the PSF membrane was also widely reported by other researchers to have high negative charge density especially at higher pH as was employed in this experiment. Therefore less flux decline or higher HPO fraction was rejected by PSF membrane with regard to the charge repulsion between negative functional groups of HPO and negative surface charge of PSF membrane, respectively. Besides that the results of Figure 4 also showed that the UV_{254} removal were always higher than DOC rejection for all the three fractions. This could be again due to the tendency or preference of polysulfone membrane to repulse the high aromatic compound such as HPO fraction compared to other fractions and that it was shown high by the percentage of UV removal, meanwhile the DOC is not a specific measurement of certain compound but a total carbon concentration comprised by the three fractions.

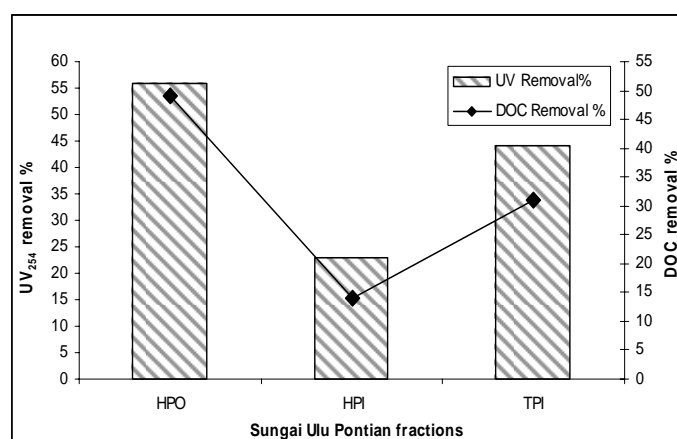


Figure 4. UV_{254} and DOC removal (%) of hydrophobic, transphilic and hydrophilic fractions by MRUTM55 membrane

The effect of membrane characteristics on NOM rejection was further studied with a MRUTM66 membrane. This membrane has lower MWCO (58 kDA), more hydrophilic and less negative charge compared to MRUTM55 membrane. From Figure 5 it was clearly shown that the MRUTM66 resulted in lesser DOC (%) rejection despite its lower MWCO than MRUTM55 membrane. This scenario is again presumably related to the shape, charge and size effects of NOM components towards the rejection mechanism of MRUTM66 membrane. In general the MRUTM66 increases the DOC rejection (%) of HPI fraction but decreases for both the HPO and TPI fractions when compared with MRUTM55.

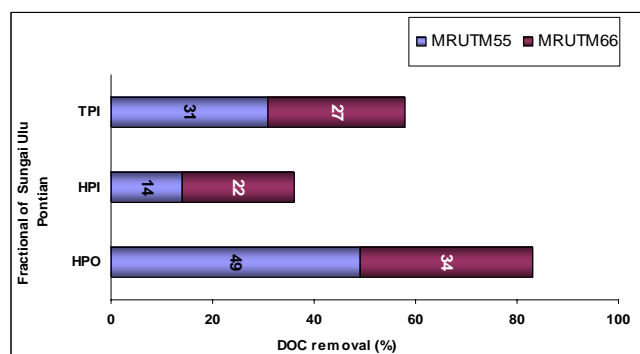


Figure 5. Comparison of DOC removal (%) between MRUTM55 and MRUTM66 on Sungai Ulu Pontian components

The HPI fraction removal (in term of % DOC rejection) by MRUTM55 was about 14% but increases to 22% when introduced to MRUTM66. However on overall the DOC removal was still dominated by HPO fraction followed by TPI and HPI fractions. Better removal in HPI fraction could be explained by the lower MWCO (50 kDa) employed by the MRUTM66 membrane whereas higher rejection of HPO and TPI than HPI component was due to higher intrinsic molecular weight (MW) possessed by both the fractions. Both the HPO and TPI fractions had been widely reported to have greater MW or higher relative molecular mass than HPI fraction by previous researchers. Furthermore the HPO fraction and TPI fraction have more complex molecules structures (aromatic shape) than HPI fraction (aliphatic structure) that made them easier to be excluded by the MRUTM66 membrane. During filtration with MRUTM66 most of the HPO fraction was successfully passing through the pores as the membrane was negatively less in charge and the electrostatic interaction between the functional groups of HPO and MRUTM66 (cellulose acetate) was no longer the rejection dominant mechanism. Thus it is worth to note here that the rejection mechanism imposed by the MRUTM66 was a steric exclusion hindrance while the MRUTM55 was an electrostatic repulsion mechanism.

3.2 NOM Fractions on Resistance in Series

The flux decline by NOM fractions were further quantitatively studied through resistance in series model (Equation 1). The explanation of flux decline caused by Sungai Ulu Pontian fractions was made by correlating their flux decline with the incurred fouling mechanisms and hydraulic resistances (Table 4). There seems to appear a relationship between the high fouling potential of HPI fraction and its high irreversible resistance (R_a). The HPI fraction has the highest R_a value (30.9%) and exhibited the worst flux decline in among the NOM fractions. Thus this severe flux decline was mostly attributed by the adsorption fouling mechanism of non-humic fraction (HPI) occurred during the filtration. This adsorption fouling mechanism can be described by saturation and precipitation processes of dissolved organic in the membrane pores; in this case it can be compounds derived from polysaccharide (neutral) and amino sugar.

Table 4. Percentage of fouling resistance to the total resistance (R_T) by NOM fractional components of Sungai Ulu Pontian

Fraction (%)	R_m (%)	R_{cp} (%)	R_c (%)	R_a
HPO	59	22	3	16
HPI	55	1.08	12.5	30.9
TPI	65	5.34	19	7

This HPI fraction was expected to have the capability in passing through the membrane pores (HPI has an aliphatic structure) and not being able to be excluded by the steric hindrance mechanism as it was smaller than membrane MWCO (68 kDa). The normal molecular mass of HPI fraction found in river waters were reported [4] to be lower than 10 kDa. In addition, as the polysaccharide does not have any ionizable functional group would definitely help them to easily overcome the electrostatic repulsion by the membrane. The precipitated HPI compounds would then adsorb onto the pores, constrict, reduce the pore sizes and to the extent of blocking it. This initial adsorption enhanced further mass accumulation of HPI that cover the pores and later lead to cake layer formation. This statement is well supported from the result of Table 4 that showed HPI fraction to have high cake resistance (12.5%). Meanwhile, the HPO fraction which has an aromatic structure showed the concentration polarization (CP) as the primary fouling mechanism to its flux declination (Table 5). This is supported by the highest R_{cp} (22%) possessed by HPO component compared to other fractions. Moreover there seems to be an underlying rejection mechanism that governs the preferential rejection of the HPO fraction by the PSF membrane that consequently led to high R_{cp} .

Table 5. Dominance fouling mechanism in NOM fractions as a function of hydraulic fouling resistance

Fraction	Primary fouling mechanism
HPO	Concentration Polarization
HPI	Adsorption
TPI	Cake formation

This result also indicates that the hydrophobicity interaction between HPO fraction and PSF membrane were found less significant in affecting the adsorptive fouling although previous studies [6] claimed that intrinsically hydrophobic membrane tend to adsorb NOM and foul more than hydrophilic membrane. This phenomenon could be well explained from the high humic content and negative charge (due to carboxylic and phenolic moieties) of HPO fraction that resulted in a

significant rejection by the negative charge of polysulfone membrane. Therefore it can be hypothesized that there are two dominance factors which are responsible in determining the HPO adsorption and type of occurring fouling mechanism. The factors are adsorptive behaviour (due to hydrophobicity property) and the electrostatic repulsion interaction (due to similarity in charge with the membrane). HPO fraction possessed negative charges functional groups that prevent it from adhering or adsorbing onto the negatively charge membrane surface and therefore impacted in high Rcp. However in this experiment the HPO fraction did exhibit a significant adsorption fouling mechanism with adsorption resistance (Ra) 16% higher than TPI fraction. The Ra of HPO fraction showed higher value than TPI fraction (7%) despite its electrostatic interaction with the membrane. This scenario was believed to happen when the hydrophobic interactions was able to overcome the electrostatic repulsion incurred by the pressure gradient of suction process. Thus for the HPO fraction, the adsorption mechanism could be taken as the secondary influential fouling mechanism after the concentration polarization. In general it could be summarized that fouling potential would be higher for the negative charge membrane when it is exposed to hydrophilic components compared to highly negative charge materials (HPO fraction) although they possessed high adsorptive tendency. The responsible foulant that affect the membrane performance was the HPI fraction which was found to performance the worst flux decline during membrane filtration. Furthermore the adsorption fouling can be taken as the primary fouling mechanism that governed the membrane flux decline as the HPI fraction had the highest Ra. Therefore in future application, NOM hydrophobicity or aromaticity (feed source characteristic) and charge interaction (membrane property) can be employed as a quantitative predictor for NOM rejection and flux declination of membrane filtration of surface water.

4.0 Conclusion

Based on the results of this experiment several conclusions can be drawn out as follows:

The HPI and HPO fraction performed strong irreversible fouling due to their neutral property and hydrophobicity interaction with the membrane. Each of the Sungai Ulu Pontian NOM fraction caused substantial flux decline and the order of fouling potential based on flux decline was HPI>HPO>TPI. Furthermore, charge interaction is more influential factor than hydrophobicity and steric exclusion mechanisms in determining the rejection, fouling mechanisms and the extent of flux decline.

References

- [1] Clever, M., Jordt, F., Knauf, R., Rabiger, N., Rudebusch, N., and Scheibel, H. 2000. Process Water Production From River by Ultrafiltration and Reverse Osmosis. *Desalination*, 131: 325-336.
- [2] Fan, L., J.L. Harris, F.A. Roddick, F.A., and Booker, N.A. 2001. Influence of the Characteristics of Natural Organic Matter on the Fouling of Microfiltration Membranes. *Wat.Res.*, 35,(18): 4455-4463.
- [3] Kaiya, Y., Itoh, Y., Fujita, K., and S. Takizawa. 1996. Study on Fouling Materials in the Membrane Treatment Process for Potable Water. *Desalination*, 106 (1996): 71-77.
- [4] Hwang, C.J., Scilimenti, M.J., and Krasner, S.W. 2000. Disinfection by Product Formation Reactivities of Natural Organic Matter Fractions of Low-Humic Water. *ACS Symposium*, 76: 173-187
- [5] Lahoussine-Turcaud, V., Wiesner, M.R., and Bottero, J.Y. Fouling in Tangential-Flow Ultrafiltration. 1990. The Effect of Colloids Size and Coagulation Pretreatment. *J. Mem. Sci.*, 52 : 173.
- [6] Nilson, J.A., and DiGiano, F.A. 1996. Influence of NOM composition on nanofiltration. *J. AWWA.*, 88(5): 53-66.

The Effect of Wall Rigidity and Deformation on Lateral Earth Pressure Distribution

Lee Min Lee¹, Lo Seng Chew², Nurly Gofar³

¹ Postgraduate Student, ² Former Undergraduate Student, ³ Senior Lecturer
Faculty of Civil Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +607-5532517, +6012-8808979, E-mail: amuellee@yahoo.com

Abstract

Development and distribution of lateral earth pressure on retaining structure is strongly affected by the rigidity and deformation behavior of the structure as well as the soil structure interaction properties. The study is aimed to get an insight to the effect of wall rigidity and their movement on lateral earth pressure and its distribution on the wall. A hypothetical problem was analyzed for different wall stiffness and deformation behavior using finite element codes implemented in Plaxis 7.2 Professional. The methodology is divided into two stages: (1) the analysis on the effect of wall rigidities represented by Young's Modulus on lateral earth pressure distribution, (2) the effect of wall movement induced by applied load on the development and distribution of lateral earth pressure. The result showed that the distribution of lateral earth pressure on wall is significantly affected by the wall rigidity and the wall movement induced by the applied load. A simplified approach is proposed for estimation of the magnitude and distribution of lateral earth pressure on retaining structure.

Keywords:

Retaining Wall, Wall Rigidity, Wall Deformation, Lateral Earth Pressure, Numerical Study

1. Introduction

The lateral earth pressure and its distribution on retaining structures depend upon several factors such as: the physical characteristics of soil, time dependent shear strength characteristics of the soil, soil-structure interaction, and deformation characteristics of the material. They are also influenced by the type and the geometry of the structure which govern the mechanism of wall movement. Furthermore, any external loads applied on the wall will also influence the lateral earth pressure on the wall.

The design and analysis of earth retaining structures and the prediction of their behavior under various loading conditions expected during the design life have been extensively investigated through theoretical, analytical as well as experimental works (Lesniewska and Porbaha, 1998). However, several issues such as the effect of wall rigidity on lateral earth pressure distribution on flexible wall, the effect of wall deformation, and the assessment of stability are still a matter of debate.

Conventionally, the design of retaining structure is based on the classical Rankine method. In this method, the wall is assumed to be rigid, move as a unit, and the soil in either side of the wall is assumed to be at its limit state. Design predictions based on Rankine's theory work quite well if the wall have sufficient rigidity to avoid bending deformations and move as a unit, but not for conditions where flexible systems are employed to retain the earth (Vaziri, 1995). Furthermore, the effects of soil-structure interaction and the processes of construction of the system is not considered in this theory.

This study was performed to evaluate the effects of wall rigidity and the wall movement induced by applied load on lateral earth pressure and its distribution.

2. Numerical modelling

Non-linear finite element model, Plaxis 7.2 Professional (Plaxis B.V., 2004), was used in this study to investigate the behavior of flexible retaining walls at working stresses and to predict wall and ground movements. Plaxis 7.2 Professional is a finite element program for geotechnical applications in which soil models are used to simulate the soil behavior.

* Corresponding Author. E-mail: amuellee@yahoo.com,
Tel: +60-7-5532517, +6012-8808979

The arrangement of the wall structure to be simulated in this study is shown in the Figure 1. The wall is simulated to retain sandy soil of 3 m high. The depth of the embedment was determined based on the Fixed Earth Method by applying factor of safety equal to one for both passive force and embedment depth. The reason behind this is to ensure that the mode of failure on the retaining wall can be determined in the study.

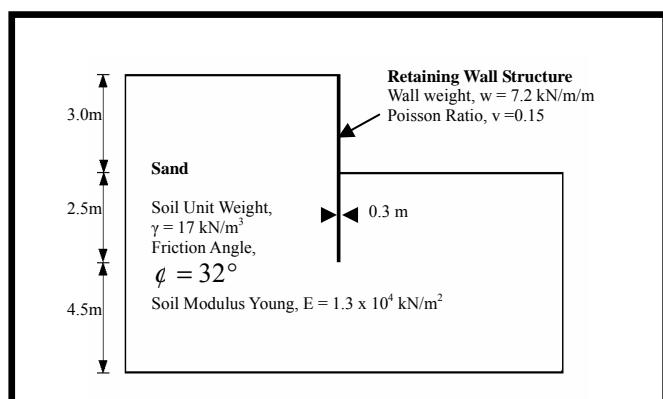


Figure 1: Arrangement of the model

The study consists of two main stages: (1) the analysis on the effect of wall rigidity represented by Young's Modulus on lateral earth pressure distribution, and (2) the effect of wall movement induced by applied load on the development and distribution of lateral earth pressure.

In the first stage of study, the retaining wall of the same dimensions was assigned with variable Modulus Young, which was increased gradually from 1.20×10^5 to 2.00×10^9 . The range was chosen based on the reality that most of the retaining walls in practice are constructed by concrete and steel with Modulus Young between 2.4×10^7 kN/m² (concrete) and to 2.0×10^8 kN/m² (steel). The variation of Modulus Young used for the study is shown in Table 1.

Table 1: Details of models simulated in Stage 1

Case no.	E (kN/m ²)	EI (kNm ² /m)	Remarks
Case 1	1.20×10^5	2.700×10^2	E failure
Case 2	1.00×10^7	2.250×10^4	
Case 3	2.40×10^7	5.400×10^4	E concrete
Case 4	4.60×10^7	1.035×10^5	
Case 5	6.80×10^7	1.530×10^5	
Case 6	9.00×10^7	2.025×10^5	
Case 7	1.12×10^8	2.520×10^5	
Case 8	1.34×10^8	3.015×10^5	
Case 9	1.56×10^8	3.510×10^5	
Case 10	1.78×10^8	4.005×10^5	
Case 11	2.00×10^8	4.500×10^5	E steel
Case 12	2.00×10^9	4.500×10^6	

For the second stage of analysis, the wall was assigned with Young's Modulus of 1.8×10^8 kN/m². The wall was assumed to move in two modes: rotation about the fixed bottom end and rotation about the fixed top end. This condition was simulated as if the wall was flexible and allowed for displacement at the top of wall only. The second case was simulated for the condition of deformation for rigid wall, in which the bottom end of the wall was allowed to rotate against the fixed top end (Clayton et.al., 1993). The movement was induced by application of surcharge on the ground surface. An additional lateral force acted on the top end of the wall was assigned to limit the displacement which may occur.

3. Results

3.1. Effect of Wall Rigidity

Results from twelve cases in the first stage showed that lateral earth pressure distributions for active condition computed by incorporating rigidity parameter do not vary much from that calculated by Rankine's theory. This shows that the soil behind the retaining structure has achieved active condition. Figure 2 shows the relationship between wall rigidity and the maximum active lateral earth pressure along the exposed wall height. The figure shows that the more rigid the wall, the higher the maximum active lateral earth pressure is. The effect of wall rigidity on these lateral earth pressure values were significant for flexible wall, and the impact reduced gradually as the wall became more rigid.

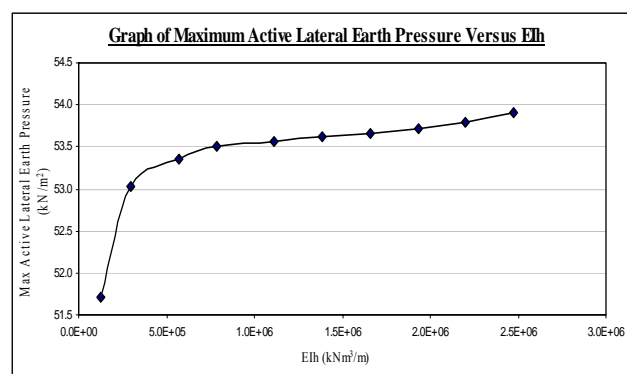


Figure 2: Maximum active lateral earth pressure versus rigidity

Large differences were observed between the results of numerical simulation and that calculated by Rankine's theory for passive pressure along the wall embedment. The maximum passive lateral earth pressure, passive thrust force and even the passive rotation moment obtained by Plaxis 7.2 Professional is actually less than predicted by Rankine's theory. At the base of the wall, all cases gave same magnitude of lateral earth pressure for both active and passive sides. This value was proven to be the magnitude of lateral earth pressure at rest. Figure 3 and Figure 4 shows the relationship between wall rigidity with the maximum passive and minimum lateral earth pressure respectively.

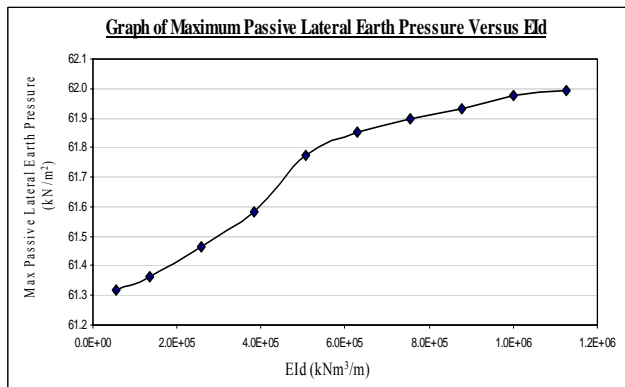


Figure 3: Maximum passive lateral earth pressure versus rigidity

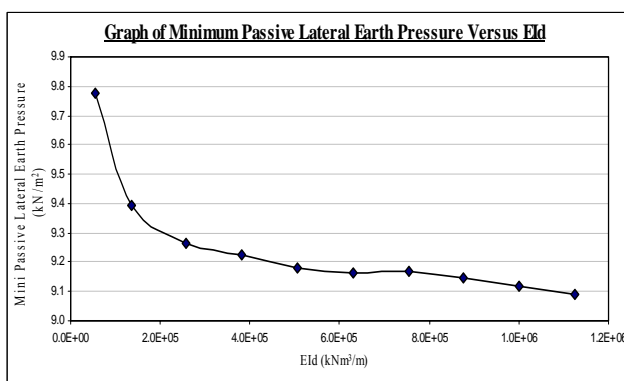


Figure 4: Minimum passive lateral earth pressure versus rigidity

3.2. Effect of Wall Movements

For the first case of analysis, free-end was assigned at the top of wall while the bottom was fixed. Movement was induced by applying a uniform load at the wall surface. Analysis shows that the maximum surcharge that the model could sustain was 10.7 kN/m^2 giving the maximum displacement at point R (Figure 5) of $6.92 \times 10^{-3} \text{ m}$ which is equivalent to $0.0023H$. This result complies with the minimum distance to achieve active condition given in literatures $0.002H$ (e.g.: Clayton et.al., 1993).

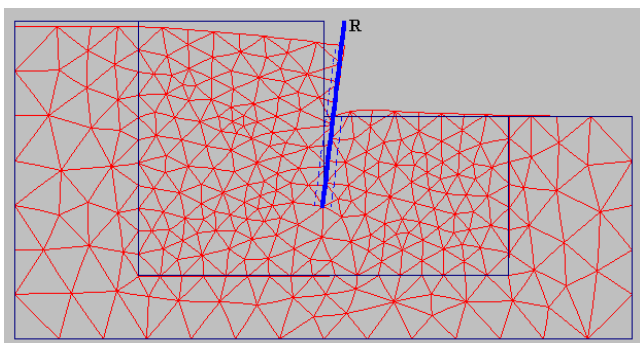


Figure 5: Wall Deformation for rotation around the bottom

Figure 3 shows the maximum horizontal displacement as the function of surcharge. The displacement expanded exponentially in which during early phase, great increment of surcharge induced a small magnitude of displacement only. This was due to the elasticity of the concrete sheet pile. However, after the sheet pile had gone through the yielding point of Point S, the displacement reacted in the other way round where little increment of surcharge increased significantly the magnitude of displacement until failure.

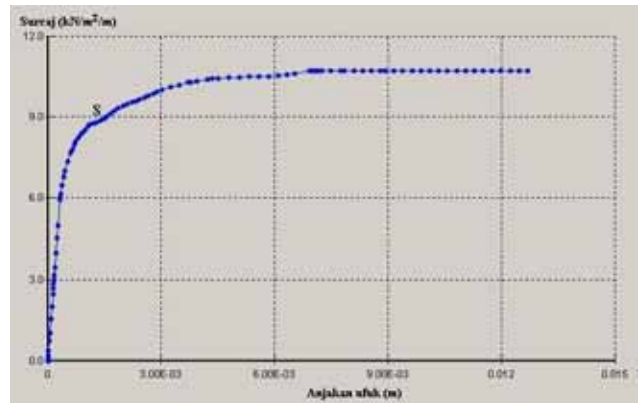


Figure 6: Deformation curve for rotation around the bottom

For the second case where fixed-end was set at the top and free-end at the bottom of the wall, the maximum surcharge recorded was 2.9 kN/m^2 , and it induced a maximum displacement of $1.23 \times 10^{-3} \text{ m}$ to the bottom end of wall as indicated by Point X in Figure 7. This maximum displacement is equivalent to $0.0004H$ ($H = 3\text{m}$) which does not comply with theoretical minimum distance to achieve passive condition given in the literature which is $0.01H$. This shows that in this case, the wall never achieve passive condition.

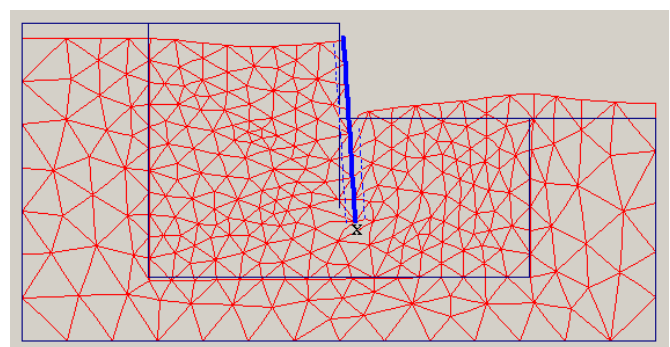


Figure 7: Wall Deformation for rotation around the top

The displacement curve for case 2 is illustrated in Figure 8. It can be seen from the curve that the magnitude of displacement increased slightly when the surcharge multiplied with a relatively greater magnitude. This was because the sheet pile was still able to withstand pressure acted behind as its strength has not reached the limit. At the yielding point of Point Y, the pattern of curve changed to a

condition where by little increment of surcharge caused great displacement. This was due to the fact that the sheet pile has gone through its strength limit. Thus, it continued to move until failure occurs.

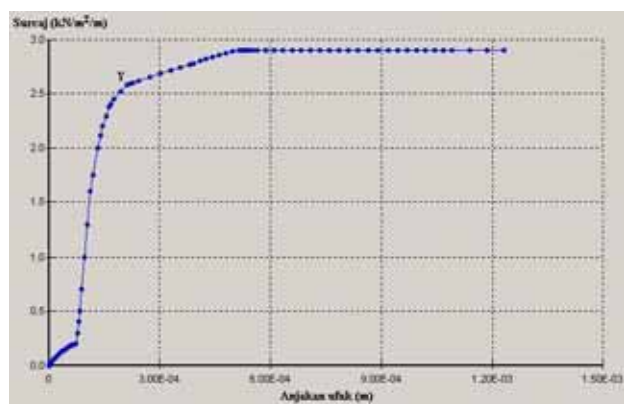


Figure 8: Deformation curve for rotation around the bottom

4. Discussion

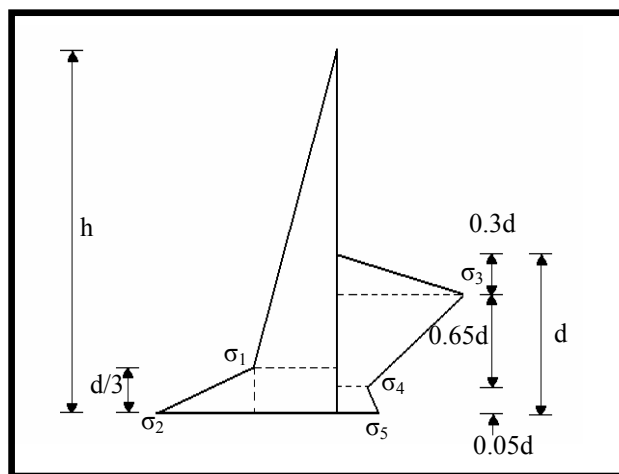
The findings have shown that the accurate prediction of lateral earth pressure and its distribution is very important in the design of retaining structures. The prediction should consider many factors mentioned in the Introduction. This study tried to include the considerations of several factors such as the property of the wall and the wall movement which represents the soil structure interaction.

The lateral earth pressure distributions in active condition observed in the study have a good agreement with that predicted by Rankine’s theory. However, the passive thrust force obtained from numerical study is half of that computed from Rankine’s. As the result, an appropriate reduction factor should be applied for when calculating the passive lateral earth pressure using Rankine’s theory as recommended in British Standard 8002 (1994).

Furthermore, the point of application of the passive thrust force predicted by Plaxis 7.2 Professional is lower than that predicted by Rankine’s theory which produces smaller bending moment. Thus, even though the adoption of the reduction factor could improve the result of lateral earth pressure from Rankine’s theory, it does not satisfy the computation of maximum bending moment.

The study also shows that the wall displacement is influenced by the wall rigidity ($\delta = f(E,I)$). The flexible wall induced greater displacement as compared to the rigid wall. Besides, the arching effect on the flexible wall caused non-linearity to the lateral earth pressure distribution on the wall (Powrie, 1997). The deformation of the wall was closely related to the lateral earth pressure acting on the wall. When the wall displaced outward from the retained soil, the active lateral earth pressure decreases and the passive lateral earth pressure increases, and vice versa.

Modification of the lateral earth pressure distribution computed from Rankine’s theory to account for wall rigidity effect is proposed based on the results of this study (Figure 6).



Note: $\sigma_1 = \gamma \times (h-d/3) \times K_a$; $\sigma_2 =$ Refer Figure 2
 $\sigma_3 =$ Refer Figure 3 ; $\sigma_4 =$ Refer Figure 4
 $\sigma_5 = \gamma \times d \times K_0$

Figure 9: Proposed simplified diagram of lateral earth pressure distribution

Result from the analysis in Stage 1 showed that flexible wall allowed for greater displacement with lower maximum bending moment. The flexible wall tends to move by rotation at its base, while rigid wall tend to move as a unit. Thus, two cases in Stage 2 were designed to simulate both conditions.

The study shows that for the case of rotation around the base, the active stage was reached at the maximum displacement of 0.002 H which complies with the theoretical minimum values suggested in literatures. This also complies with the findings by Goh (1993) that the retaining wall tend to deflect more at the top than near the bottom because of rotation effect resulting from the combination of horizontal and vertical movements.

As surcharge applied on the soil surface, it induced pressure to the soil and the soil distributed the pressure on the surface of the wall. In order to resist the pressure induced on the back of sheet pile, the top end of the wall moved away from the backfill. The circumstance caused displacement at the top of the wall. The factor of safety (FOS) decreased gradually as the surcharge increased, and it failed (FOS<1.0) when the wall displacement exceeded the limit.

On the other hand, the case of rotation around the top did not give reasonable results because the length of embedment inhibits the movement to reach passive condition. Never the less, the maximum surcharge observed from this study is much less than that obtained for case 1 whereby the wall is assumed to rotate about its base.

5. Conclusion

Based on the present study, it can be concluded that the wall rigidity affect the lateral earth pressure acting on wall. The more rigid the wall, the higher the maximum active lateral earth pressure is. The effect is significant for flexible wall, and the impact reduced gradually as the wall became more rigid. Besides, rigid wall tends to displace as a mass unit while flexible wall tends to displace by rotation at its base.

The deformation of wall influences the lateral earth pressure distribution and the factor of safety of the wall. As the wall displaces, the lateral earth pressure acting on the wall increases either in active or passive condition depends on the direction of the wall movement. Good agreement was achieved with regard to the active condition, but much less movement was observed for passive condition in term of pressure distribution.

Theoretically, greater displacement is required to achieve passive condition as compared to active condition, provided that the wall initially is in at rest condition. This study showed that the maximum displacement was achieved when the wall failed in active condition. On the other hand, much less displacement was observed when the wall failed in passive condition which might be due to the effect of embedment.

In respect of factor of safety (FOS), regardless of which direction and how the wall deforms, the FOS will decrease gradually as the wall displace from its original location and fails when it reach its limit.

6. References

- [1] Clayton, C.R.I., J. Milititsky, R.I. Woods, *Earth Pressure and Earth Retaining Structure*, 2nd ed. Chapman and Hall, 1993.
- [2] Goh, Anthony T.C. Behaviour of Cantilever Retaining Walls, *Journal of Geotechnical Engineering*, ASCE Vol. 119, No. 11: 1751-1770
- [3] Lesniewska Danuta A., Porbaha Ali. 1998. *Numerical simulation of scaled retaining walls by rigid-plastic approach*. Computers and Geotechnics 23: 113-129.
- [4] Plaxis BV. 2004. PLAXIS Finite Element Code for Soil and Rock Analyses. Netherlands.
- [5] Powrie, William. 1997. *Soil Mechanics : Concepts and Applications*. London: E & FN Spon.
- [6] Vaziri H.H. 1996. A Simple Numerical Model for Analysis of Propped Embedded Retaining Walls. *Int. J. Solids Structures*. Vol. 33, No. 16: 2357-2376.

Effect of Binder Sand Ratio to Basic Properties of Epoxy Grout

S. K. Lim¹, F. Zakaria²,
M. W. Hussin³, Z. Abd. Hamid⁴, R. Muhamad⁵

^{1,2,3} Faculty of Civil Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5532158, Fax: +60-7-5531967, E-mail: siongkang@hotmail.com

^{4,5} Construction Industry Development Board (CIDB), Malaysia
Grand Season Avenue, 53000 Kuala Lumpur, Malaysia

Abstract

Nowadays resin based grout is becoming one of the important structural repairs materials. Low viscosity epoxy resin is used in the formulation of epoxy grouts in this study up to its characteristics such as high compressive strength, good adhesion to the substrate, chemical resistant etc. The viscosity and compressive strength of epoxy grouts is significant to determine their suitability as structural cracks repair materials. This paper discussed the effect of binder sand ratio to compressive strength development of epoxy grout. The design of epoxy grouts should consider the economic factor, pumpability (viscosity) and achieving of optimum strength above 70MPa under air curing condition. Hardener to resin ratio used in this study is vary from one part of hardener to three to six parts of epoxy by weight. The dry graded fine sand was added to resin matrix vary from 6 to 9 parts by weight depend on consistence flow of epoxy grout. This study found that epoxy grout with binder (hardener and epoxy) sand ratio 6:10-12 part by weight provided optimum compressive strength, consistent flow and sufficient working time for structural crack repairs.

Key words: Epoxy grout, compressive strength, viscosity, working time

1. Introduction

Resin grouts are widely used to seal structural cracks, restore the integrity of the structure and prevent the ingress of water, which would cause corrode any metal reinforcement. Since the early 1960s the use of resin compositions in the construction industry has grown from very small beginnings to significant tonnages due to its bond strength that are considerably greater than the cohesive strength of concrete [1]. The rate of development of mechanical strength and time saving is varying depend upon the resin system used, the ambient temperature and other factors. The cost in repairing and strengthening faulty structures makes it essential that the materials used will provide the durability to match the residual design of the structure. Resin based grout offer real independence in term of its engineering properties and performance when use in structural crack repairs.

1.1. Epoxy Resin

Epoxy resins, also known as epoxide or ethoxyline resins; contain the epoxy group, which is the chief centre of their reactivity. When manufactured, the resins are either liquids or solids and contain, on average, two epoxy groups per molecule. In this physical state, the resins are thermoplastic, that is, they can be repeatedly soften heat and hardened by cooling. The essential feature of this epoxy resin technology

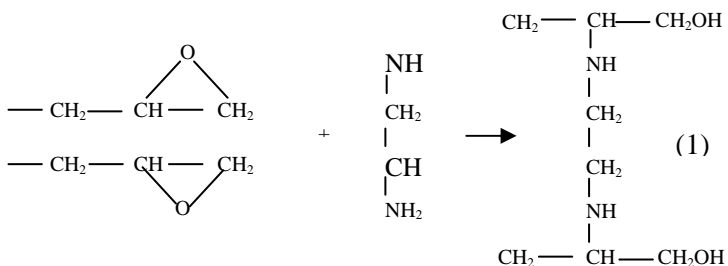
is the conversion into a hard, infusible three-dimensional network in which the resin molecules are cross-linked together by means of strong covalent bonds. In fact, some epoxy resins are almost pure organic chemicals [2]. The most common and widely used epoxy resins are made by allowing polycondensation involving epichlorohydrin and bisphenol A. For example, the epoxy diglycidyl ether of bisphenol A, DGEBA [3].

Epoxy resin possesses higher strength than conventional repair materials along with controlled shrinkage. The compressive strength is the highest compare to different repair materials. In general, epoxy resin showed the highest compressive strength at all ages, whereas the conventional repair materials gave the lowest values. Epoxy resin posses superior tensile strength, which is 1/5 of its compressive strength value and about four times greater than conventional repair materials [4]. Meanwhile, epoxy resin based grout provides excellent strength and bond to concrete, with low shrinkage and permeability [5].

Over the past 40 years the chemistry of the epoxy resin components used in most construction applications has not changed significantly. Change has been made to improve the storage stability of epoxy at low temperature and to reduce potential health hazards that have been identified. There have, however, been major changes in the chemistry of the hardener systems, mainly to improve the short and long term technical performance of the resin systems.

1.2. CURING AGENT/ HARDENER

Epoxy resins are cured by using catalysts or cross-linking agents such as amine, anhydride and acid. The amine functionality reacts relatively with the epoxide group to give well-cured resins at room temperature. There are two major groups of amines, aliphatic amines and aromatic amines. The aliphatic amines are widely used for curing epoxy resins which can react at room temperature to cured products from 24 – 48 hour while aromatic amines can give cured epoxy materials of good chemical and thermal resistance, are not cure conveniently at room temperature. The classic epoxy curing mechanism is the reaction between a primary amine and an epoxy functional group. The two cross-linking reactions are external, by reaction of the oxirane function group with active hydrogen, and internal, by homopolymerization through the oxirane oxygen [6]. The epoxy rings at the ends of the two linear molecules can be reacted with amine like ethylenediamine to form the cross-link, according to the reaction scheme shown below. The OH groups that are formed act as reaction sites for further cross-linking reactions.



Equation (1) shows the epoxy rings at the ends of the two linear molecules reacted with ethylenediamine to form the cross-link

2. Experimental Program

2.1. Raw Materials Selection

The raw materials that used to produce epoxy grout as discussed in this paper are described below:

- Low viscosity epoxy resin – Epoxy resin is a high cost polymer material. The common epoxy resin is formed by epichlorohydrin and bisphenol A, or known as diglycidyl epichlorohydrin of bisphenol A (DGEBA). This research convert using modified DGEBA resin known as low viscosity epoxy resin as raw material. The monofunctional reactive diluents especially monofunctional glycidyl ethers, are used at relatively low percentages to achieve

lower viscosities and form part of the polymer backbone of DGEBA resins, to impart a measure of flexibility without the possibility of migration. The low viscosity resin used is complying with requirement of Table 2; Type I, Class A epoxy resin according to ASTM D 1763 – 81[7]. The epoxy resin typical properties are show in Table1. The chemical structure of DGEBA resin is show in Equation (2).

- Hardener – Type of hardener used to cure low viscosity DGEBA resin is based on manufacturer' suggestion. It is reaction products of polyamine with fatty acids like linoleic acid, linolenic acid and arachidonic acid, each one has carboxyl function group (COOH) to produce amidopolyamines or amides provide the largest group of commercial hardeners for adhesive applications. Basically, when carboxylic acids react with ammonia, its form ammonium salts. Ammonium salts are ionic substances when dissolved in water or heated. When heated, ammonium salts of carboxylic acids are converted to amides as showed in Equation (3) [8].
- Sand – The type of sand used is standard washed sand according to ASTM C 778 – 91[9]. Fine aggregate was oven dried at 105°C about 24 hours to mitigate moisture content of sand so that no affected the polymerization process of resin matrix. The totally dried sand then sieved to remove litter/ rubbish and graded according to the requirement of this study.



Fig. 1: Binder and graded dry sand

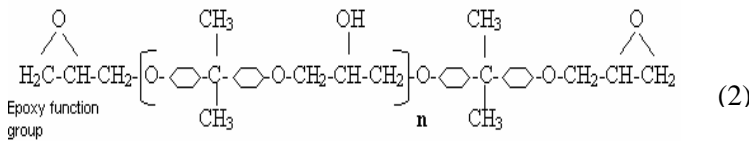
Table 1: Typical properties of low viscosity modified DGEBA epoxy resin with hardener (Ref. manufacturer's specifications)

Resin : Hardener ratio (by volume)	4 : 1
Specific gravity	1.09
Viscosity (Brookfield) Spindle #3/ 20rpm, 20 ⁰ C	250cP
Tensile strength (N/mm ²) BS6319: Part 7 at 20 ⁰ C	20
Compressive strength (N/mm ²) BS 6319: Part 2 at 20 ⁰ C – 7 days	75
Flexural strength (N/mm ²) BS 6319 Part 3 at 20 ⁰ C	55
Tensile bond strength (N/mm ²) Dry concrete at 20 ⁰ C Wet concrete at 20 ⁰ C	Concrete failure in all cases
Modulus of elasticity (N/mm ²) in flexure	4500
Useable pot life at 20 ⁰ C	40-60 minutes

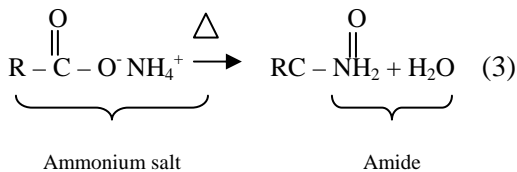
	EP-CTR			
	A1	B1	C1	D1
Binder (epoxy resin + hardener) to Sand Ratio	4 to 6	5 to 9	6 to 11	7 to 13
Flowable (Brookfield Viscometer)	Consistent flow (1500 to 2000 centipoises)			
Gel Time (minute) (ASTM D 2471)	35 - 50			

2.3. Mixing Procedure

First, graded dry sand was weighed and introduced to clean and dry container. Proposed amount of epoxy resin was weighed and its hardener was added to epoxy in other clean and dry container. The resin and its hardener are thoroughly mixed. Then, the dry sand was added to the resin-hardener system. The binder and sand was thoroughly mixed about 2 to 3 minutes. Certain amount of fresh mixed resin grout is tested for its viscosity/ flowable and working/ pot life and the balance was poured into the 50mm x 50mm x 50mm mould before gel form obtained. The moulds used for casting resin matrix were made by steel or timber as shown in Fig. 2 and Fig. 3. Before casting, the moulds must be applied with special demoulding wax to ease the demoulding works. After one day, the samples were demould and subjected to air curing in laboratory, which the average temperature is 30⁰C with 65% relative humidity.



Equation (2) is the chemical structure of DGEBA epoxy resin



Equation (3) shows the converted of ammonium salt of carboxylic acid to amide

2.2. Mix Proportion

Table 2 presents the mix designs of the epoxy grout in this study. The binder sand ratio of epoxy grouts were determined by viscosity and consistent flow of grout.

Table 2: Details mix proportions for epoxy grouts

By weight	
-----------	--



Fig. 2: Steel mould for casting cube



Fig. 3: Timber mould for casting cube

2.4. Test Methods

The compressive test was performed according to ASTM C 579 – 01[10] and equivalent to BS 6319: Part 2[11]. 50mm x 50mm x 50mm cube has been selected as the specimen size. Specimens were tested at a rate from 1.5 to 1.8kN/s according to ASTM C 579 – 01. Three cube test specimens were prepared for each period and curing condition and results are the mean of individual results. The gel time of epoxy grouts measured in accordance to ASTM D 2471[12]. After mixing, the epoxy composite was poured into the clean paper cup or similar container. Record the start of mixing as the “starting time.” Every 15 seconds, probe the center of the reacting mass, with the wooden stick perpendicular to the material surface. When the reacting material no longer adheres to the end of a wooden stick, record the “gel time” as the elapsed time from the start of mixing. The flowability/viscosity of the epoxy grouts was tested with Brookfield rotational viscometer at spindle 3, 60rpm, 30°C. The measurement of grout matrix by rotational viscometer was shown in Fig. 4.

3. Results and Discussion

The results for compressive strength of polymer samples prepared with low viscosity epoxy resin, hardener and dry graded sand, which subjected to air curing regime, are presented and discussed. The viscosity and gel time of epoxy matrix with different binder to sand ratio also discussed in this study.

The following codes were used to present the results:

EP = samples prepared with epoxy resin.

Alphabet (A1, B1, C1, D1) = Different binder to sand ratio based on total mass of grout mixes (refer to Table 2).

3.1. Flowability and Gel Time

When the epoxy used in this study was tested with rotational viscometer at spindle 3, 60rpm, 30°C, the reading obtained

was 120 to 130 centipoises. When amide (hardener) that has very low viscosity was added, the resin system was diluted and has much lower viscosity. The low viscosity of resin grout can ease the pumping or injection works. It was found in this study that adding of graded sand with binder sand ratio according to Table 2 into epoxy resin matrix maintained the consistent flow of the grout and the viscosity of grout mixes were in the range of 1500 to 2100 centipoises at spindle 3, 60rpm, 30°C which consider as low viscosity bonding systems based on ASTM C 881[13]. When the binder sand ratio decreased or sand content increased in resin matrix, the flowability of the grouts was decreased when tested with rotational viscometer. This is because the fine graded sand in resin matrix made its shear rate higher. When the shear rate higher, the viscous of the grout mixes increased.



Fig. 4: Measure of resin grout's viscosity by Brookfield rotational viscometer

The gel time of the epoxy resin mixture mainly control by the epoxy to hardener ratio. When higher content of hardener was added, the resin matrix was cured faster. However, the minimum epoxy to hardener ratio is around two to one to cure the system and made it solid. If the hardener is more than resin, the resin matrix will not go through the polymerization process to make it solid. In the other hand, the very high epoxy to hardener ratio also made the systems cannot completely polymerize. The percentage of hardener was determined by the objectives to have sufficient working time and maintain the strength required. It was also found that when adding more graded sand, the increasing of viscosity of grout mixes was slightly shorten the gel time or working time of resin matrix. Table 2 showed the gel time of epoxy grouts measured in accordance to ASTM D 2471. The results in Table 2 indicated the mix proportions in this study have sufficient working time that fulfill the requirement of resin bonding system in ASTM C881.

3.2. Compressive Strength

Compressive strength was determined according to ASTM C 579. Three 50 x 50 x 50 mm cubes were weighed and tested for load at uniform rate 1.7kN/s until failed. Fig. 5 presented the average compressive strength for each batch of epoxy grout under air curing conditions at 1, 7 and 28 days. Fig. 6 and showed the compressive test of grout cubes and Fig. 7 showed the failure mode after compressive tests.

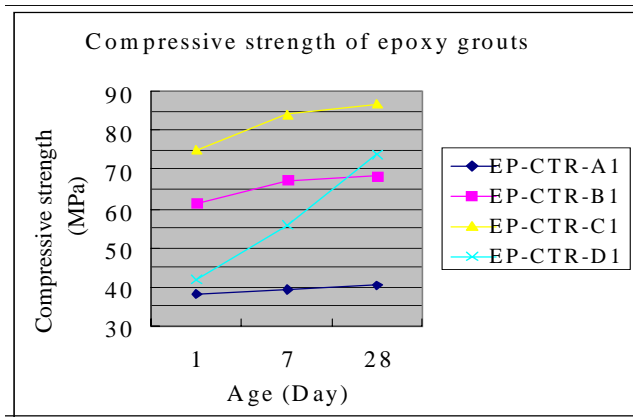


Fig 5: Compressive strength of epoxy grout under air curing condition



Fig 6: Compressive strength test



Fig. 7: Failure mode of epoxy grout cubes after compressive test

From Fig. 5, we can notice that different binder sand ratio of epoxy grouts was obtained little different compressive strength by ages. EP-CTR-A1 almost obtained the optimum value of compressive strength at day-1 and maintained its strength after that. The compressive strength of EP-CTR-A1 at day-28 was around 40MPa. Although the graded sand added to EP-CTR-A1 was the least among the mixes, but the epoxy to hardener ratio was not compatible as the hardener concentration was too much to cure the epoxy caused the polymerization process was very fast and almost completed within 1 day period. The percentage of cross-linked by epoxy matrix was not enough to provide the high strength to the mix. EP-CTR-B1 achieved the strength above 60MPa at day-1 and maintained its strength after day-7 around 65MPa to 68MPa. Although the binder to sand ratio of EP-CTR-B1 was lower than EP-CTR-A1, but the compressive strength of EP-CTR-B1 was higher than EP-CTR-A1. This is because the hardener to epoxy ratio of EP-CTR-B1 was more compatible compare with EP-CTR-A1 to cure and increased the cross-linked of epoxy matrix and also bonding between the epoxy matrix and aggregate. Although the higher content of aggregate added to polymer matrix can reduce the strength of the mix, but the effect was little. The most importance function of aggregate or filler added to polymer matrix was reduced cost by increased the overall volume of mix, reduced high exothermic of the mix to prevent thermal contraction, maintained stability etc. EP-Ctr-C1 provided the highest compressive strength among the mixes as its binder to sand ratio was the most compatible to able the highest percentage of polymerization (cross-linked). EP-CTR-C1 achieved the compressive strength more than 85MPa at day-28. EP-CTR-D1 has the lowest binder to sand ratio and achieved the strength around 75MPa at day-28. The trend of strength development of EP-CTR-D1 was slower and different compared with other mixes. The other three mixes almost fully cured and achieved the optimum strength at day-7 but the EP-CTR-D1 still have high strength development after day-7 until day-28. The strength was increased from 55MPa at day-7 to 75MPa at day-28 or increased its strength around 36% within the period. The hardener to epoxy ratio of EP-CTR-D1 was the lowest among the mixes, the lower concentration of the hardener caused the curing process of polymer matrix slower and the strength development also need to take longer time compared with other three mixes. The optimum strength of the EP-CTR-D1 was the second highest among the mixes at day-28.

4. Conclusion

Based on the-above investigations, we can conclude the fine aggregate content added into the epoxy matrix has little effect to the compressive strength of epoxy grout. The higher the content of sand inside the mix, the compressive strength

of cube reduced slightly. The viscosity of the epoxy grout with lower binder to sand ratio was higher as the higher content of sand increased the shear of the overall mix. Based on the above results, the epoxy grout can maintain its strength and durable under air condition in tropical climate. The strength development of the epoxy matrix mainly depends on the percentage of its polymerization (cross-linked). The compatible hardener to epoxy resin ratio play the importance role to enhance the cross-linked between the functional group (monomer) of polymer matrix and the bonding to the aggregate. According to the ASTM C 881, the requirement of the adhesive system of resin matrix that can bond efficiently to hardened surface is the compressive strength at day-7 should achieved 69MPa or above. Among the mixes of this study, EP-CTR-C1 fulfilled the requirement of adhesive system in ASTM C 881.

Acknowledgements

This research was carried out in the Structure and Material Laboratory of Faculty of Civil Engineering, Universiti Teknologi Malaysia. Financial support from the Construction Industry Development Board of Malaysia (CIDB) is greatly appreciated.

References

- [1] Shaw J. D. N. 1985. Resins in Construction. Cement Composites and Lightweight Concrete. *Proceedings of the Seminar Resins in Construction*, London, 7(4): 217-223.
- [2] Potter W. G. 1975. *Uses of Epoxy Resins*. London.: Newnes-Butterworths.
- [3] Premamoy G. 1996. *Polymer Science and Technology of Plastic and Rubbers*. New-Delhi.: Tata McGraw-Hill Publishing Company Limited. 287-290
- [4] Hassan K. E. 2000. Effect of hot-dry curing environment on the intrinsic properties of repair materials. *Cement and Concrete Research* 22(6): 453-458.
- [5] Saccani A. 1998. Durability of epoxy resin-based materials for the repair of damaged cementitious composites. *Cement and Concrete Research* 29(7): 95-98.
- [6] Goulding T. M. 1994. Epoxy Resin Adhesives. *Handbook of Adhesive Technology*. New York: Marcel Dekker Inc.1-13, 531-546.
- [7] American Society for Testing and Materials 1988. Standard Specification for Epoxy Resins (ASTM D 1763 - 81) (Re-approved 1988).
- [8] Morris H., Leo R. B., Scott P. and Susan A. 1993. *Introduction to Organic and Biochemistry*. Pacific Grove.: California. Brooks/ Cole Publishing Company.
- [9] American Society for Testing and Materials (1991). Standard Specification for Standard Sand (ASTM C 778 - 91).
- [10] American Society for Testing and Materials (2001). Standard Test Method for Compressive Strength of Chemical-Resistant Mortars, Grouts, Monolithic Surfacing, and Polymer Concretes (ASTM C 579 - 01).
- [11] British Standard (1983). Method for Measurement of Compressive Strength. Testing of Resin Compositions for Use in Construction (BS 6319: Part 2)
- [12] American Society for Testing and Materials (1988). Standard Test Method for Gel Time and Peak Exothermic Temperature of Reacting Thermosetting Resins (ASTM D 2471 - 88).
- [13] American Society for Testing and Materials (2002). Standard Specification for Epoxy-Resin-Base Bonding Systems for Concrete (ASTM C 881 - 90).

The Effect Of Palm Oil Fuel Ash Content And Water Dry Mix Ratio on The Density And Compressive Strength Of POFA Cement Based Aerated Concrete

K. Abdullah¹, M. W. Hussin²,
F. Zakaria³, Z. Abd. Hamid⁴, R. Muhamad⁵

^{1,2,3} Faculty of Civil Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5532158, Fax: +60-7-5531967, E-mail: khairunisa1978@yahoo.com

^{4,5} Construction Industry Development Board (CIDB), Malaysia
Grand Season Avenue, 53000 Kuala Lumpur, Malaysia

Abstract

The 21st century which well-known as era of globalization has indeed lead to active transfer of knowledge and technology all over the world resulting in more innovations towards existing product by researchers everywhere in order to accommodate consumer demand. After 10 years Faculty of Civil Engineering of Universiti Teknologi Malaysia successfully revealed the use of palm oil fuel ash (POFA) as partial cement replacement in normal concrete. Now again during the era of information and technology, the same faculty again discovered that the abundantly produced by-product of palm oil mill has the potential to be incorporated as a partial cement replacement in aerated concrete. This new revelation has lead to more efforts in studying the properties of this material in aerated concrete. Therefore, the current objective of this investigation is to study the effects of increasing the levels of POFA replacement and varying the water-dry mix ratio towards the density and compressive strength of aerated concrete. Experiments have been conducted by replacing 10%, 20%, 30%, 40% and 50% of POFA by weight of Ordinary Portland Cement and also varying the water dry mix ratio used in the range of 0.2 to 0.3. All specimens were cast in the cubes of (70.6mmx70.6mmx70.6mm) and then subjected to water curing. The compressive strength of the specimens was obtained by testing the cured cubes at the age of 28days. Meanwhile, the density of the specimens was determined by subjecting it to oven dry based on procedures stated in AAC 4.1. In conclusion, only appropriate water dry mix ratio and POFA content able to produce a lower density POFA aerated concrete as compared to control aerated concrete and at the same time successfully exhibits strength within the range that a normal aerated concrete would have.

Key words: palm oil fuel ash, partial cement replacement, aerated concrete, compressive strength, density

1. Modernization Of Concrete

Since the early age, human civilization has started inventing things for the purpose of survival. However as time passes by, the dynamic human mind made them easily became weary of using the same tools or method in their life and always yearn for improvement. As a result of the end user demand, until now researchers is motivated to continuously work industriously leading to design of new material or modification of the existing things to be better than before and at the same time fulfill the current consumer requirement. In the field of building industry, construction material specifically concrete is one of the areas that have been subjected to constant research in order to support the growing industry need besides offering more choices of material for contractors to be used in construction project. As a result, now there is various type of concrete with unique names each possessing special feature that

successfully fulfilling current construction need are available in the market. Nowadays, aerated concrete is begin to gain contractors interest due its two special attribute that is lightness and versatility as compared to normal concrete. Documented history acknowledges that Sweden is the first country to produce aerated concrete commercially in 1929 (Short & Kinniburgh, 1978). Then within the period of 75 years, this in-depth studied concrete has been produced in various types and methods for the use of construction in many countries.

Generally, building constructed using aerated concrete is lighter as compared to the similar one constructed with normal concrete. Either researchers Short & Kinniburgh (1978) and Narayanan & Ramamurthy (2000b) in 20th or 21st century, claimed that aerated concrete which is light contribute to the reduction of building dead weight thus resulting in more economic structural design. In addition, the lightness of the aerated concrete structure also makes it easier to be transported and handled (Short & Kinniburgh,

1978 and Holt & Raivio, 2005). The low density of this concrete makes it only require few machineries and labors for hauling and handling process that definitely decrease the level of difficulties in work besides increase contractors profitability. Besides that, aerated concrete also possesses few characteristic of a timber in which it is also easy to work with because it can be sawn, chiseled, planed, screwed and nailed (Short & Kinniburgh, 1978 and Holt & Raivio, 2005). Conclusively, aerated concrete offers more flexibility to the contractors when it is able to provide them cost and time saving through easier modifying process as compared to existing normal concrete.

2. Aerated Concrete In Current Trade

Since the earliest of 20th century, aerated concrete has been added as one of the alternative construction material to support the building trade. The continuously developing aerated concrete research gained several names throughout the years has such as cellular concrete, cellular aerated concrete, gas concrete or foamed concrete. Structurally aerated concrete is not dense and heavy as the existing normal concrete. Instead it has a lot of voids evenly formed within the internal structure of this light concrete. Each void is divided by the sections that been created by sand that been bind by cement through hydration process. Basically, aerated concrete is either a cement or lime mortar, classified as lightweight concrete, in which air-voids are entrapped in the mortar matrix by means of a suitable aerating agent (Narayanan & Ramamurthy, 2000b). Gas concrete which is also known as aerated concrete among the researchers can be produced by using various types of gas forming materials that will react inside the slurry mortar to produce air voids.

Narayanan and Ramamurthy (2000b) elaborated about the reaction of aerating agents used by highlighting that aluminium powder, hydrogen peroxide / bleaching powder and calcium carbide liberate hydrogen, oxygen and acetylene, respectively. The most common method is addition of Aluminium powder which is added to the mixing ingredients at about 0.2% to 0.5% by dry weight of cement (Holt & Raivio, 2005). Following that, another researcher Grutzeck et.al, (2004) explained in detail that the cellular character is obtained by adding Aluminium powder to the mixture at the very last minute, pouring slurried mixture into a mold and then allowing it rise much like a cake rises in the oven. After the "cake" gains enough "green strength" to allow it to be demolded, the cake is cut into desired shapes using gangs of wire saws (Grutzeck et.al, 2004).

Basically, slurry mixture was acquired by firstly producing a dry mix consisting fine aggregates, cement and pore-forming chemical with final addition of sufficient water followed by fast mixing process until uniform mix is obtained instantly. Then, the pore-chemical reacts with the presence of water producing air-voids within the aqueous mix at atmospheric pressure making the mix expand to certain extent depending on the amount of gas produced and

entrapped inside the structure. Once the concrete stopped expanding and set, then the overflowed concrete was trimmed using a sharp flat edge to acquire a uniform surface.

2.1. Strength And Density of Aerated Concrete

Gas concrete which is also known as aerated concrete is favored by the contractors due its special feature that is being lighter than the existing concrete. Shrivastava (1977), a researcher in the early days claimed that this concrete can be manufactured in a wide range of densities to suit specific requirement. However, Narayanan & Ramamurthy (2000b) one of the researcher in millennium year managed to specifically mention that by appropriate method of production, aerated concrete with a wide range of densities (300-1800 kg/m³) can be obtained thereby offering flexibility in manufacturing products for specific application (structural, partition and insulation grades. The same researcher admitted that aerated concrete is characterized by the presence of large voids deliberately included in its matrix to reduce the density. In other words, the more voids confined in the hardened aerated concrete so the more porous the concrete structure will be thus leading to a lower density of the concrete. Meanwhile, compressive strength is one of the important properties that need to be considered in order to determine its function in building construction. Generally, the compressive strength for aerated concrete is ranging from 2 to 5 N/mm² (Taylor, 2000). Studies done by Holt & Raivio (2005) stated that aerated concrete has lower strength compared to standard concrete. The absence of coarse aggregate as one of the ingredient in aerated concrete is among the factors that lead to lower strength as compared to standard concrete. The presence only fine sand bonded together with cement and well distributed pores available within the concrete structure able to produce a lower density concrete. However, Narayanan & Ramamurthy (2000b) concluded that compressive strength of aerated concrete increases linearly with density but increase in moisture content of aerated concrete will cause decrease in its compressive strength. In general, Narayanan & Ramamurthy (2000b) also added that composition and method of curing influence the physical and mechanical properties of aerated con

3. Innovation Of Aerated Concrete

Actively advancing communication throughout the world has results in active knowledge sharing among researchers due to existence of good communication technology has lead Although, there are a large number of studies done and papers published on aerated lightweight concrete (Tam et.al, 1987; Hauser et.al, 1999; Narayanan & Ramamurthy, 2000a; Goual et.al, 2000; Kearsley & Wainright, 2002; Holt & Raivio, 2005; Jones & McCarthy, 2005) but until now there's no study done on aerated concrete produced using agricultural ash particularly POFA as partial cement

replacement material. Faculty of Civil Engineering from Universiti Teknologi Malaysia has become the pioneer in Malaysia to venture on research of gas concrete which is a type of aerated concrete. Earlier Mat Yahaya (2001) studied aerated concrete that was incorporated with agricultural ash and then followed by another researcher (Arreshvhina, 2002) who designed a new design mix of gas concrete known as slag cement based aerated lightweight concrete in Universiti Teknologi Malaysia.

Currently, attempt has been made to produce aerated concrete known as palm oil fuel ash cement based lightweight aerated concrete. Accomplishment in incorporating POFA as partial cement replacement material in producing new type of lightweight concrete could reduce amount of cement used as compared to ordinary aerated concrete. Moreover, usage of POFA in this product could also prevent the abundantly produced by-product by Malaysian palm oil industry from causing significant pollution to the environment thus fashioning cleaner surrounding for healthier lifestyle to the community surrounding besides assisting this industry to be more environmental friendly and at the same time provide opportunity to increase their profit making.

3.1 POFA And Chemical Composition

Palm oil fuel ash is a by-product produced in palm oil mill. 10 years ago it was dumped as waste as reported by Sumadi & Hussin (1995) and until now the situation remained unchanged. The ash produced sometimes varies in tone of colour from whitish grey to darker shade based on the carbon content in it. Basically, the physical characteristic of POFA is very much influenced by the operating system in palm oil factory. Normally, after combustion about 5% palm oil fuel ash by weight of solid wastes is produced (Sata et.al, 2004). Since, Malaysia is continuously paving the way towards increasing palm oil production, therefore it is expected more ashes will be produced and failure to find any solution in making use of this by-product will create severe environmental problems that in turn diminish the level of healthy lifestyle lead by community surrounding.

Both physical properties and chemical analysis indicated that POFA is a pozzolanic material (Awal & Hussin, 1997; Sumadi & Hussin, 1993). This pozzolanic material is grouped in between Class C and Class F as specified in ASTM C618-92a (Awal & Hussin, 1997). POFA is moderately rich in silica content meanwhile lime content is very low as compared to OPC (Awal & Hussin, 1997). However, the POFA chemical composition can be varied due to operating system in palm oil mill.

3.2 POFA Along The Discovery Path

Initiative that been taken since 1990s to study POFA has finally lead to unique discoveries benefiting the palm oil industry and community. Until now, some researchers (Hussin & Awal, 1996; Awal & Hussin, 1997; Sata et.al,

2004) that has been attentively studying on POFA use has successfully reveal the positive impact of POFA in concrete technology. The revealed facts are more in the terms of enhancement towards the properties of concrete either strength or durability aspect. It has been discovered that POFA concrete gain maximum strength when 30% of the cement was replaced with POFA (Hussin & Awal, 1996 and Awal & Hussin, 1996). However, further increase in the ash content would reduce the strength of concrete gradually (Awal & Hussin, 1996). Studies done also lead to conclusion that increasing in fineness of POFA would lead to greater concrete strength development than the coarser one (Hussin & Abdul Awal, 1996).

Meanwhile, the durability performance of the concrete consisting POFA is also better than normal concrete. Findings have also revealed that incorporation of POFA as partial cement replacement in concrete result in significant increase in chemical resistance to acidic environment (Awal & Hussin, 1997). POFA possess good characteristic towards chemical attack especially sulphate and acid and also other chemical agent when it is used as partial cement replacement in concrete mix Awal (1998). Moreover, despite the higher alkali content POFA, it has been effective in suppressing expansion due alkali-silica reaction. However, performance of POFA when incorporated as partial cement replacement in aerated concrete still remains undiscovered.

4. Experimental Programme

4.1. Material Preparation

4.1.1 Palm Oil Fuel Ash

In this present study, POFA was collected from a factory processing palm oil owned by Yayasan Pembangunan Johor in Ladang Alaf at the area of Bukit Lawang, Johor Darul Takzim. The ash was found at the flue of tower where all the fine ashes that were trapped while escaping from the burning chamber of the boiler. Among the available ashes there, only the one looks greyish were sorted out and collected. Firstly, collected POFA were dried in the oven at the temperature of $110^{\circ}\text{C} \pm 5$ for 24 hours in order to remove moisture in it.

Secondly, the dried ashes were sieved through a $300\mu\text{m}$ sieve in order to remove bigger size of ash particles and impurities. Only the fine ashes passing through $300\mu\text{m}$ sieve were collected and ground in a modified Los Angeles abrasion test machine having 10 stainless bars which each of it is 12 mm diameter and 800mm long in order to acquire finer particles. The fineness of the ash was checked by wet sieving through $45\mu\text{m}$ sieve at every half-hour grinding of 4kg ash. Finally the grounded ashes were stored in airtight container and kept in a humidity-controlled room to isolate from the atmospheric humidity. Table 1.0 shows the chemical composition of palm oil fuel ash.

4.1.2 Cement

A single batch of Ordinary Portland Cement (OPC) which is produced by HOLCIM under the brand name of "SELADANG" was used throughout the experiments. The OPC to be used throughout this research was stored in airtight container in order to maintain the quality of cement. Table 1.0 shows the chemical composition of Ordinary Portland Cement (OPC).

4.1.3 Fine Aggregate

In order to produce POFA cement based aerated concrete only fine sand was used. The sand obtained was dried in the oven at the temperature of $110^{\circ}\text{C} \pm 5$ for 24 hours in order to remove moisture in it. Then, the oven dry sand was sieved passing the 600um sieve before it was stored in an airtight container from atmospheric humidity.

4.1.4 Aluminium Powder

Aluminium powder was used as the gas-forming agent to create air voids in the POFA cement based aerated concrete produced. Table 2 shows the specification and chemical composition of the Aluminium powder used.

4.1.5 Superplasticizer

A chemical admixture was used in preparing POFA cement based aerated concrete in order to decrease water content and to achieve early high strength. The superplasticizer incorporated in this research is in a form of dry powder known as Sulfonated Naphthalene Formaldehyde Condensate

4.2. Manufacture And Compressive Strength Testing

4.2.1 Mix Proportion

In order to study effect of POFA content towards the density and strength of POFA cement based aerated concrete, 6 mixes of OPC/POFA concrete containing various proportions of POFA were cast and cured for 7 and 28 days. POFA was mixed as weight-for-weight replacement cement, the replacement levels being 10%, 20%, 30%, 40% and 50%. A normal aerated concrete mix containing 100% OPC was used as control specimen. All specimens were produced by adding with required sand, cement, Aluminium powder, Superplasticizer and adequate water dry mix ratio and cast in form of cubes (70.6x70.6x70.6mm).

Meanwhile, effort to study effect of water dry mix ratio towards POFA cement based aerated concrete was conducted by preparing 6 mixes of OPC/POFA concrete containing various proportions of POFA. Each mixes was used to prepare 3 types of specimen containing different water dry mix ratio but with a constant use amount of sand, POFA, cement, Aluminium powder and Superplasticizer. All

specimens were cast in form of cubes (70.6x70.6x70.6mm) and cured for 7 and 28 days.

4.2.3. Process Of Studying Compressive Strength And Density

For compressive strength 70.7mm cubes were cast, three specimens were being tested for a particular mix. All freshly cast specimens were left in the formwork for 24 hours before being demolded and then subjected to water curing until it is time to be tested. Compressive test were conducted at the age of 7 and 28 days. The optimum pozzolana content will be determined base on the strength achievement on 28 days. According to Massazza (1993), the optimum pozzolana content refers to the 28day strength. All methods following sampling, making and testing of specimens were in accordance with BS1881: Part 116. The density of all specimens was taken by drying the specimen in the oven for 24 hours based on the procedure enlisted in AAC 4.1.

Table 1: Properties of Palm Oil Fuel Ash And Ordinary Portland Cement

Chemical Constituents	POFA	OPC
Silicon Dioxide (SiO_2)	53.82	28.2
Aluminium Oxide (Al_2O_3)	5.66	4.9
Ferric Oxide (Fe_2O_3)	4.54	2.5
Calcium Oxide (CaO)	4.24	50.4
Magnesium Oxide (MgO)	3.19	3.1
Sodium Oxide (Na_2O)	0.1	0.2
Potassium Oxide (K_2O)	4.47	0.4
Sulphur Trioxide (SO_3)		2.3
Phosphorus Oxide (P_2O_5)	3.01	<0.9
Loss On Ignition (LOI)	10.49	2.4

Table 2: Properties of Aluminium Powder

Properties	
Specification	No.300
Color	Silver
Particle Size	Mesh 220
Chemical Composition	%
Aluminium	Min 99.3
Copper	Max 0.1
Iron	Max 0.4
Silica	Max 0.2

5. Results and Discussion

5.1. Investigation Of Strength And Density (Effect Of POFA)

It is mentioned earlier that concrete were cast with OPC and OPC with POFA added at various percentage replacement level. Here, all the specimens were tested for compression at the age of 28 days. Table 3 reveals that there is strong relationship between the strength and density of aerated concrete. All specimens exhibit high strength when the density is also high and vice versa. This condition is clearly acceptable since Narayanan & Ramamurthy (2000b) clearly indicated that compressive strength of aerated concrete increase linearly with density. Although it can be observed, increase in the POFA replacement level result in decrease of both compressive strength and density but it is still to early to conclude that POFA could not gives higher strength than normal aerated concrete. This is because Narayanan & Ramamurthy (2000b) has clearly stated that the strength of Non Autoclaved Aerated Concrete, NAAC increases 30 – 80% between 28 days and 6 months and marginally beyond this period. Therefore, the researcher firmly believe that there will be suitable percentage of POFA that can be used to partially replace the cement used and at the same time improve the strength of aerated concrete. Thus, conclusion to be made on the exact percentage of POFA replacement for achievement of optimum strength of POFA aerated concrete still yet to be reserved due to the facts the researcher wish to observe the strength development for long term period.

Table 3: Compressive Strength Performance And Density Of Aerated Concrete Without And With Various Level % Of POFA Replacement At 28 days

% POFA Replacement	Strength (MPa)	Density (kg/m ³)
0	7.70	1055
10	6.10	1008
20	5.84	983
30	3.81	904
40	1.95	845
50	0.96	731

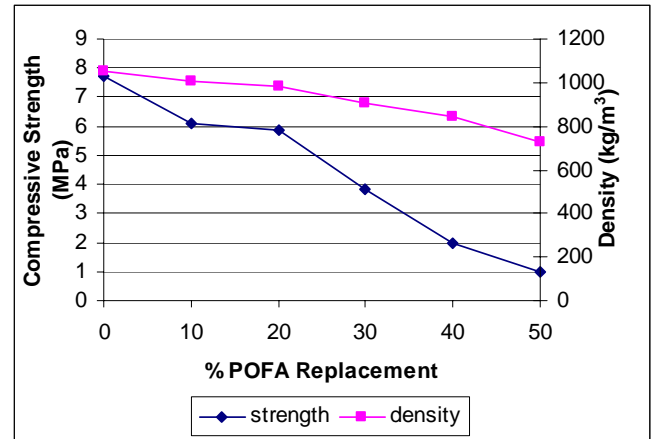


Fig 1: Effect Of POFA Replacement Towards The Compressive Strength And Density Of Aerated Concrete With Various Level Of POFA Replacement

5.2. Investigation Of Strength And Density (Effect Of Water Dry Mix Ratio)

In the case of specimens prepared by varying water dry mix ratio for each mix consisting constant ingredients, all specimens were subjected to water curing before tested for 7 and 28 days. Based on Table 5, it can be seen clearly that decrease in water dry mix ratio definitely increase both density and compressive strength of all specimens. Less water added not only make the process of mixing become more difficult but also retard the aeration process. This is because the Aluminium powder used could not have enough water to react and become hydrogen gas forming bubbles with the mix. Hence, incomplete aeration causes less expansion of the mortar mix.

Meanwhile, too much of water dry mix ratio able to provide a very low density of POFA cement based aerated concrete but with low strength. Excessive water used definitely assists in mixing process when mix becomes very slurry. However, on the other hand too much of water cause voids rupture when it have to flow out from the mix when expansion happen. Inadvertently, the void that is supposed to be entrapping hydrogen gas was ripped open by the flowing out water eventually causing the gas released and decreases the density of hardened paste. This finding is also supported by Narayanan & Ramamurthy (2000) stated that increase in moisture content of aerated concrete will cause decrease in compressive strength of aerated concrete.

Table 4: Compressive Strength Performance And Density Of Aerated Concrete Containing Various Level % of POFA Replacement With each Mix Added With 3 Different Water Dry Mix (WDM) Ratio At 28 days

% POFA	Water dry mix ratio			Strength	Density
	Low	Exact	High		
0	0.205	0.210		8.00	1100
				7.70	1055
			0.220	6.10	1040
10	0.210	0.212		6.80	1032
				6.10	1008
			0.214	5.70	967
20	0.212	0.214		6.00	1028
				5.84	983
			0.216	4.11	917
30	0.214	0.216		4.21	1005
				3.81	904
			0.218	2.89	880
40	0.216	0.218		2.25	942
				1.95	845
			0.220	1.37	837
50	0.218	0.220		1.20	876
				0.96	790
			0.222	0.40	731

Fig 2: Effect Of 3 Different Water Dry Mix Ratio (Low, Exact, High) Towards The Compressive Strength Of Each Aerated Concrete Mix Consisting Varies POFA Replacement Level

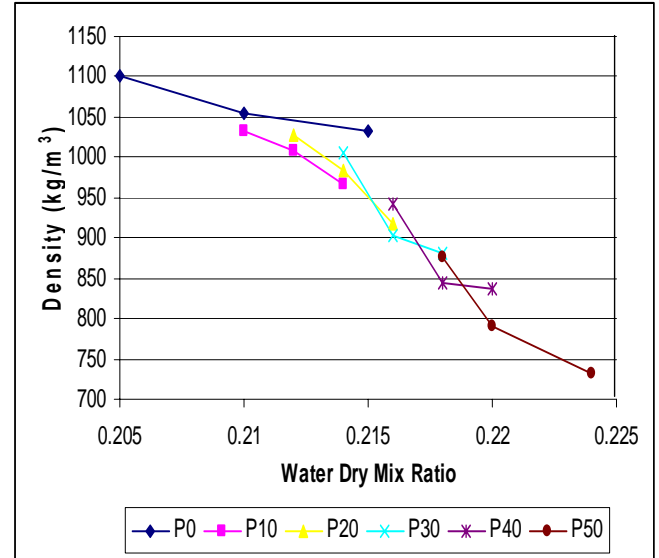


Fig 3: Effect Of 3 Different Water Dry Mix Ratio (Low, Exact, High) Towards The Density Of Each Aerated Concrete Mix Consisting Varies POFA Replacement Level

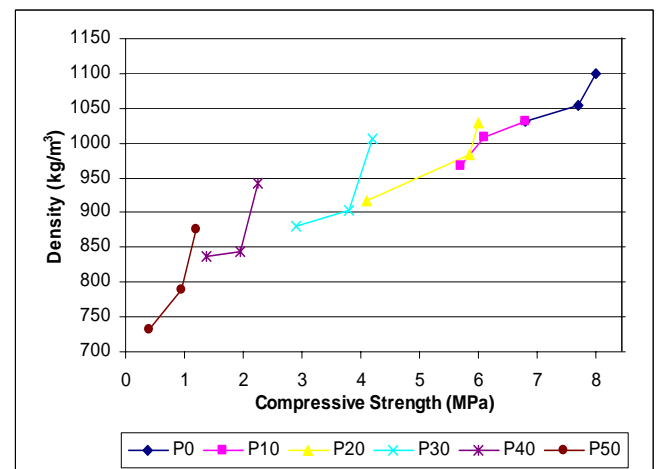
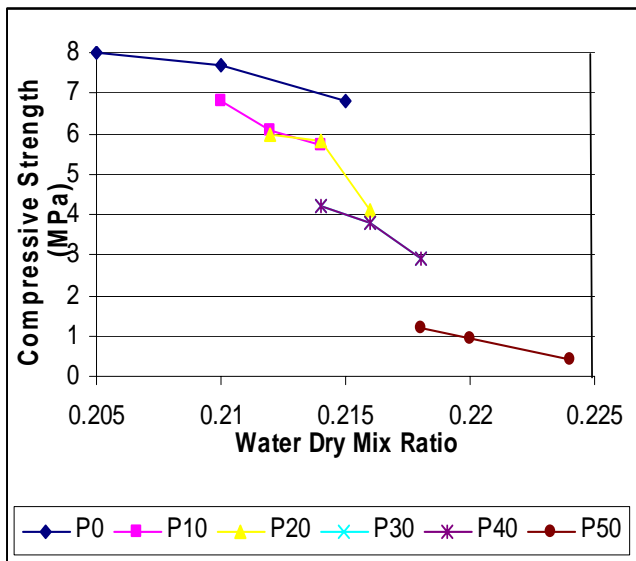


Fig 4: Effect Of 3 Different Water Dry Mix Ratio (Low, Exact, High) Towards The Compressive Strength And Density Of Each Aerated Concrete Mix Consisting Varies POFA Replacement Level



6. Conclusion

Basically, the results obtained for both studies indicated that the compressive strength and density of POFA cement based

aerated concrete are influenced by water dry mix ratio and POFA content. This finding is also aligned with statement by Narayanan & Ramamurthy (2000) whom highlighted that water content and characteristics of ingredients used is among the factors influence the compressive strength of aerated concrete. Generally, proper trial mix need to be carried out to determine adequate water content needed for complete aeration purpose and exact percentage of POFA to be replaced in order to acquire the targeted strength and density of aerated concrete. This study could make the contractors understand the behavior of aerated concrete if exact water dry mix ratio and POFA content failed to be used.

Accomplishment in determining appropriate amount of water content for the exact percentage POFA that could give optimum strength for aerated concrete with lower density compared to existing aerated concrete definitely able to introduce new concrete material for Malaysian construction industry. Creation of new mix introducing the use of POFA in aerated concrete is hoped to open new horizon for palm oil industry to settle issue of this problematic by-product and become more environmental friendly industry in future.

Acknowledgements

This research was carried out at the Structure and Material Laboratory of Faculty of Civil Engineering, Universiti Teknologi Malaysia. Financial support from the Construction Industry Development Board of Malaysia (CIDB) is greatly appreciated. The researcher also wishes to express her gratitude towards Kolej Universiti Kejuruteraan Dan Teknologi Malaysia (KUKTEM) for sponsoring her current studies.

References

- [1] Short & Kinniburgh. 1978. *Lightweight Concrete*. 3rd Edition Applied Science Publisher Limited
- [2] Taylor, G.D. 2000. *Materials In Construction: An Introduction*. 3rd Edition. Pearson Education Limited.
- [3] Narayanan,N & Ramamurthy,K. 2000b. Structure And Properties Of Aerated Concrete : A Review *Cement & Concrete Composite*. (22) 321-329
- [4] Holt, E & Raivio,P 2005. Use Of Gasification Residues in Autoclaved Aerated Concrete. *Cement & Concrete Research*(16) 623-628
- [5] Narayanan,N & Ramamurthy,K. 2000b. Microstructural Investigations On Aerated Concrete. *Cement & Concrete Research*. (30)
- [6] Tam, C.T. Lim,T.Y. Ravindarajah, R.S. and Lee, S.L. 1987. Relationship Between Strength And Volumetric Composition Of Moist-Cured Cell Aerated Concrete. *Magazine Of Concrete Research* Vol.39 (138) 12-18.
- [7] Hauser, A. Eggenberger, U. and Mumthaler 1999. Fly Ash From Cellulose Industry As Secondary Raw Material In Autoclaved Aerated Concrete. *Cement & Concrete Composite*. (22) 297-302
- [8] Goual, M.S. 2000. Estimation Of The Capillary Transport Of Clayey Aerated Concrete Using A Gravimetric Technique. *Cement & Concrete Research* (34) 1559-1563
- [9] Kearsley, E.P. and Wainwright, P.J. 2001. Porosity And Permeability Of Foamed Concrete. *Cement & Concrete Research* (31) 805-812
- [10] Jones,M.R. and McCarthy,A. 2005. Utilising Unprocessed Low-Lime Coal Fly Ash In Foamed Concrete. *Fuel* (84) 1389-1409
- [11] Sumadi, S.R. and Hussin, M.W. 1995. Palm Oil Fuel Ash (POFA) As A Future Partial Cement Replacement In Housing Construction. *Journal Of Ferrocement* (25) 25-34
- [12] Sata, Jaturapitakkul, C. and Kiattikomol, K. 2004. Utilization Of Palm Oil Fuel Ash In High-Strength Concrete. *Journal Of Materials In Civil Engineering, ASCE* (16) 623-628
- [13] Awal, A.S.M and Hussin, M.W 1997. Some Aspects Of Durability Performances Of Concrete Incorporating Palm Oil Fuel Ash. *Proceedings Of Fifth International Conference On Structural Failure, Durability and Retrofitting*. Singapore 210-217
- [14] Sumadi, S.R. and Hussin, M.W. 1993. Agricultural Ash (AA) – Construction Material For The Future. *Kongres Sains & Teknologi Malaysia* (1) 105-114
- [15] Nagataki,S. 1994. Mineral Admixtures In Concrete : State Of The Arts And Trends. *Proceedings Of The V.Mohan Malhotra Symposium On Concrete Technology: Past, Present and Future*, ACI Publication Sp-144 447-482
- [16] Awal, A.S.M 1998. A Study Of Strength And Durability Performances Of Concrete Containing Palm Oil Fuel Ash. Universiti Teknologi Malaysia. PhD Thesis

- [17]Abu, A.S.M 1990. The Pozzolanicity Of Some Agricultural Fly Ash And Their Use In Cement Mortar And Concrete. Universiti Teknologi Malaysia. Master Thesis
- [18]Arreshvhina,N. 2002. Application Of Slag Cement Based Aerated Lightweight Concrete In Non-Load Bearing Wall Panels. Universiti Teknologi Malaysia. Master Thesis
- [19]Mat Yahaya,F 2003. Pengaruh Abu Terbang Kelapa Sawit Terhadap Kekuatan Mampatan Dan Ketahananlasakan Konkrit Udara. Universiti Teknologi Malaysia. Master Thesis

Separation of Suspended Particles from Wastewater using Electrocoagulation Technology

Moh Faiqun Ni'am¹ , Fadil Othman^{2*} , Johan Sohaili² , Zulfa Fauzia²

¹ Civil Engineering Department, Faculty of Engineering
Universitas Islam Sultan Agung, Semarang 50112, Indonesia
Tel: +62-24-6583584, Fax: +62-24-6582455, E-mail: irfani_mt@yahoo.com

² Environmental Department., Faculty of Civil Engineering
Universiti Teknologi Malaysia, Johor, Malaysia
E-mail: dfadilos@utm.my , j_sohaili@yahoo.co.uk , zulfa_fauzia@yahoo.com

Abstract

This paper reports preliminary stage from a study that combined magnetic field and electrocoagulation (EC) technology for removal of suspended particles in wastewater treatment. The experiments carried out in this work were setup into static methods. Batch experiments with two monopolar iron (Fe) plate anodes and cathodes were employed as electrodes. The effects of important process variables such as current density and operating time on the chemical oxygen demand (COD) and turbidity removal efficiencies were explored. Wastewater sample was made from powder milk with concentration 700 mg/L and acidic condition was employed (pH~3). DC current was verified at 0.5 – 0.8 A, and operating time at 30 – 50 minutes. The results show that the effluent wastewater was very clear and its quality exceeded the direct discharge standard. In addition, the experimental results also show that the electrocoagulation can neutralize wastewater pH.

Keywords :

Electrocoagulation, Iron Electrode, Suspended Particles, Sedimentation.

1. Introduction

Electrocoagulation (EC) has been used for the treatment of wastewater by various authors, and several differences were found in comparison to the chemical coagulation process. A literature survey indicates that EC is an efficient treatment process for different wastes, e.g. soluble oils, liquid from the food, textile industries, or cellulose and effluents from the paper industry (Carmona et al., 2006; Kumar et al., 2004; Calvo et al., 2003; Larue et al., 2003; Holt., et al. 2002). EC is an effective process for the destabilisation of finely dispersed particles by removing hydrocarbons, greases, suspended solids and heavy metals from different types of wastewater (Carmona et al., 2006; Kumar et al., 2004). According to Can et al. (2006), EC has been proposed in recent years as an effective method to treat various wastewaters such as: landfill leachate, restaurant wastewater, salina wastewater, tar sand and oil shale wastewater, urban wastewater, laundry wastewater, nitrate and arsenic bearing wastewater, and chemical mechanical polishing wastewater.

Pletcher and Waish (1990, in Lin et al., 1998) explained that the mechanism of the Electrochemical

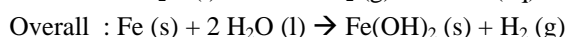
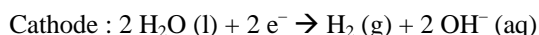
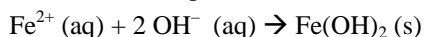
process in aqueous systems is quite complex. It is generally believed that there are three possible mechanisms involved in the process: electro-coagulation, electro-flotation and electro-oxidation. According to Can et al. (2006), three main processes occur during EC; (i) electrolytic reaction at electrode surfaces, (ii) formation of coagulants in the aqueous phase, (iii) adsorption of soluble or colloidal pollutants on coagulants, and removal by sedimentation or flotation.

Aluminium or iron are usually used as electrodes and their cations are generated by dissolution of sacrificial anodes upon the application of a direct current (Carmona et al., 2006). Kobya et al. (2003) has been investigated EC technologies to treatment of textile wastewaters using iron and aluminum electrode materials. The results show that iron is superior to aluminum as sacrificial electrode material, from COD removal efficiency and energy consumption points.

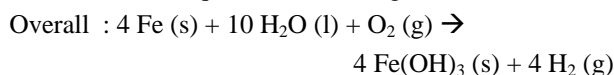
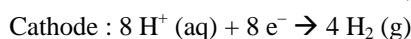
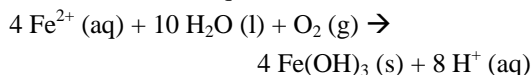
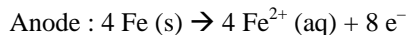
In EC with electrical current flowing between two electrodes, coagulant is generated in situ by electrolytic oxidation of the anode material. With an iron anode, $\text{Fe}(\text{OH})_n$ with $n = 2$ or 3 is formed at the anode. Simplified oxidation and reduction mechanisms at the anode and cathode of the iron electrodes are represented by (Mollah et al., 2001; Daneshvar et al., 2003; Larue et al., 2003; Daneshvar et al., 2004; Daneshvar et al., 2006) :

* Corresponding Author. E-mail: dfadilos@utm.my ,
Tel: +60-7-5531630, Fax: +60-7-5566157

(a) Mechanism 1 :



(b) Mechanism 2 :



According to Larue et al. (2003), the generation of iron hydroxides (Fe(OH)_n) is followed by an electrophoretic concentration of colloids (usually negatively charged) in the region close to the anode. Particles interact with the iron hydroxides and are removed either by surface complexation or electrostatic attraction (Mollah et al., 2001; Daneshvar et al., 2003 – 2006).

2. Experimental Details

2.1. Wastewater Characteristics

The synthetic wastewater sample made from milk powder and treated by using HCl 1 M as pH adjustment and electrolyte. The composition of artificial wastewater is shown in Table 1. Concentration of HCl in this fluid is 5 mL/L (0.5 %) and obtained maximal current 0.8 A at voltage 30 V. The current density was adjusted to a desired value and the coagulation was started.

Table 1. Characteristics of wastewater

Parameter	Value
Chemical oxygen demand COD (mg/L)	1140
Total suspended solids TSS (mg/L)	1400
Turbidity (NTU)	491
Initial pH	7.45
pH after adjusted by HCl	2.91

2.2. Experimental Setup

The experiments carried out in this work were setup into static methods. EC and fluid wastewater batch tests were performed in a cylindrical glass cell (volume 2000 mL) with stirring at constant speed (Fig. 1). Stirring was provided by a plate impeller from plastic material (\varnothing 3 cm) at a rotating velocity of 100 rpm (HEIDOLPH RZR-2101 Electronic).

For EC tests, the monopolar iron (Fe) plate electrodes (130 x 50 x 4 mm) were used in this work and the distance between both was 50 mm. Electrodes were placed in two litres of fluid wastewater and connected to terminals of a DC Power Supply (LODESTAR 8107 ; 30V / 10A) with

potentiostatic or galvanostatic operational options. The DC current was kept constant at 0.50A, 0.65A and 0.80A. Bubbles generated from water electrolysis in EC could float flocs to the top of the suspension. After EC process, fluid wastewater were degassed under low stirring speed with an impeller velocity 30 rpm.

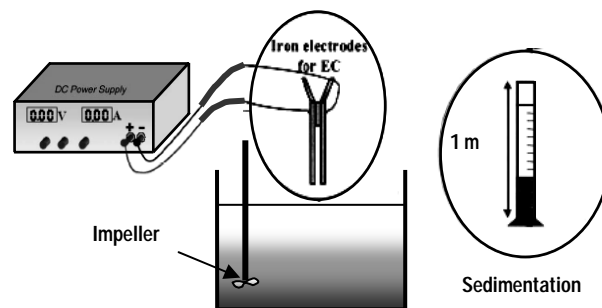


Figure 1. Static system apparatus

The fluid wastewater after EC treatment were placed in graduated sedimentation columns of 5 cm internal diameter and 1 m height (Fig. 1). During each 30 minute settling time, the effects of relevant wastewater characteristics such as pH, turbidity and suspended solid (SS) removal efficiencies have been explored.

2.3. Analytical Procedure

The pH of the solutions was measured by pH meter (Thermo Orion Model 420 A+). Thermal analysis was performed and equipped using HANNA Instrument Checktemp HI-98501. The turbidity removal was measured from wastewater samples by HACH DR/4000 (HACH Method 10047). COD measurements were determined according to the Standard Methods for Examination of Water and Wastewater (APHA, 1992). The COD samples were analysed using UV-Vis HACH DR/4000 spectrophotometer (HACH Method 8000).

To measure TSS, the wastewater samples were filtered through a standard GF/F glass fibre filter. The residue retained on the filter was dried in an oven at 103^oC to 105^oC until the weight of the filter no changes. The increase in weight of the filter represents the total suspended solids (APHA Method 2540 D).

The calculation of turbidity and suspended solid removal efficiencies after electrocoagulation treatment were performed using this formula (Daneshvar, et al., 2006) :

$$CR(\%) = \frac{C_0 - C}{C_0} \times 100 \quad (1)$$

Where C_0 and C are concentrations of wastewater before and after electrocoagulation in NTU or mg/L, respectively.

3. Results and Discussion

Floc size and density significantly influence the performance of liquid – solid separation processes. In principle, large and dense flocs are preferable since they

have high sedimentation velocities and are more easily dewatered (Larue et al., 2003).

Electrocoagulation (EC) is susceptible to produce flocs of higher size and density. EC is considered as a low sludge producing technology. The flocs formed by EC are relatively large, contain less bound water, are more stable, and therefore, amenable to filtration. Chemical aids are not required in EC which can be easily integrated with conventional waste control system (Mollah et al., 2001; Mollah et al., 2004; X. Chen et al., 2000).

3.1. Effect of EC on pH Values

In the previous studies, such as were explained by Daneshvar et al. (2006) and Kobya et al. (2006), pH is an important factor influencing on the treatment performance of the electrocoagulation process. Generally, the pH of the medium changes during the process, as observed also by other investigators (Kobya, M., et al., 2003 and X. Chen, et al., 2000).

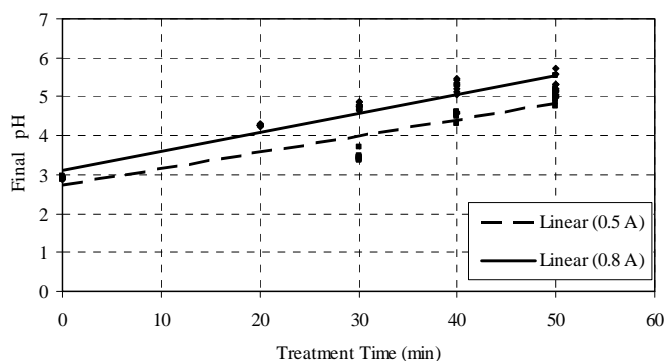


Figure 2. Effect of EC process on values change of pH

As seen clearly in Figure 2, the pH increases as the time of EC process is increased. It is happened because OH⁻ ion accumulates in aqueous solution during the process. The increase of pH at acidic condition was attributed to hydrogen evolution at cathodes.

3.2. Effect of Treatment Time

The effluent treated with iron electrode as anode, appeared greenish first and then turned yellow and turbid. This green and yellow colour must have resulted from Fe²⁺ and Fe³⁺ ions generated during the EC process. Fe²⁺ is the common ion generated in situ of electrolysis of iron electrode. It has relatively high solubility at acidic or neutral conditions and can be oxidised easily into Fe³⁺ by dissolved oxygen in water.

Fig. 3 illustrates the removal of COD and turbidity as a function of treatment time. It is clearly seen from Fig. 3 that, the treatment time has a significant effect on the pollutant removal. When the treatment time changed from 10 to 50 minute, the removal of COD from 967 to 267 mg/L and turbidity from 447 to 9 NTU were obtained.

In this process, electrocoagulation involves two stages which are destabilization and aggregation. The first stage is usually short, whereas the second stage is relatively long.

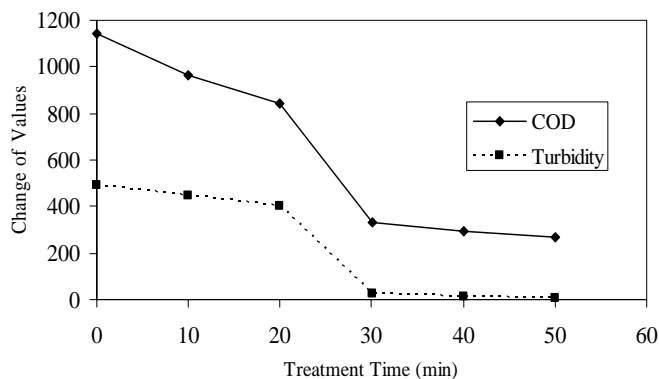


Figure 3. Effect of treatment time on the removal of COD and Turbidity (at 0.8 A current)

From Fig.3 explainable that metal ions as destabilization agent are produced at the anode through electrochemical reactions. When the treatment time was quick, so the charge loading was low. In the condition like that, the metal ion (Fe³⁺) dosage was not sufficient to destabilize all colloidal and finely suspended particles. Thus the COD and turbidity removal efficiencies were not high.

3.3. Effect of Settling Time

From the batch studies involving various settling time presented for Total Suspended Solid (TSS) removal experiment (Figure 4), the best efficiency started after 120 minutes of sedimentation. As seen in Fig. 4, Suspended particle removal is increases as the settling time is increased. With an addition of treatment time from 30 becomes 50 minutes, TSS removal is increased from 17.86% to 31.86% on 24 hours of settling time. Whereas on 12 hours of settling time, TSS removal is increased of between 15.71% – 31.29%.

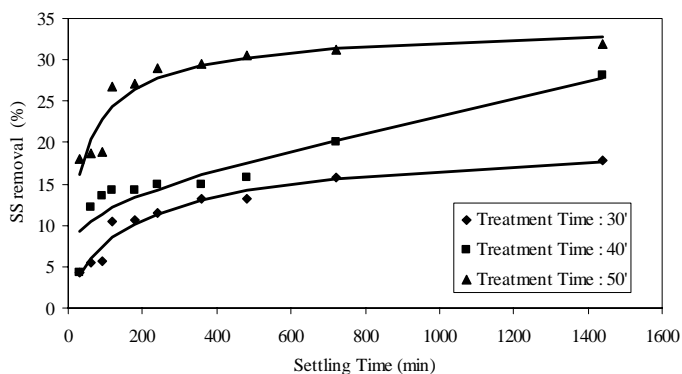


Figure 4. TSS removal efficiencies as a function of settling time (at 0.8 A current ; treatment time 30 – 40 min)

As depicted in Fig. 5, the turbidity removal efficiency is increased to a value more than 95%. On 24 hours of settling time, turbidity removal is increased to 94.50% for 30 minutes treatment time. While 40 minutes is 97.96% and 98.17% for 50 minutes treatment time.

The behaviour of those process (Fig. 4 and Fig. 5) explain that the EC technology can enhances the settling velocity of suspended particles.

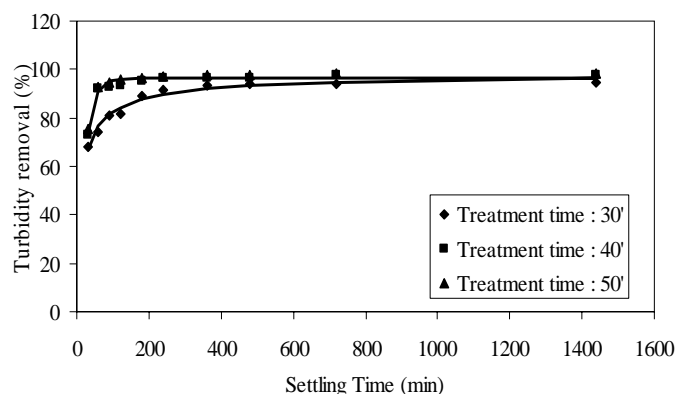


Figure 5. Turbidity removal efficiencies as a function of settling time (at 0.8 A current ; treated at 30 – 40 min)

4. Conclusion

Suspended particles removal from wastewater by EC technique was measured in the batch studies of a single stirred apparatus. The monopolar iron (Fe) plate electrodes were used in this work and were set at distance of 50 mm in the cylindrical glass cell (volume 2000 mL).

The results show that the suspended particles and turbidity removal efficiencies are as high as 31.86% and 98.17% on 24 hours of settling time. Whereas the removal of COD from 967 to 267 mg/L (76.58%).

In general, the results was obtained from the curves of treatment time and settling time, explain that the EC technology can enhances the settling velocity of suspended particles. It mean that the EC technology can separate the suspended particles from wastewater and improve its quality.

Acknowledgements

The authors are grateful for financial support of this project by the IRPA under Grant : VOT 74146, the University Technology Malaysia and Islamic University of Sultan Agung Semarang for financial and other supports provided. We special thanks to Mr. Dadan Hermawan and Ms. Salmiyati for helpful and useful suggestions.

References

- [1] Calvo, L.S., J. P. Leclerc, G. Tnguy, M. C. Cames, G. Paternotte, G. Valentin, A. Rostan, and F. Lapique. 2003. An Electrocoagulation Unit for The Purification of Soluble Oil Wastes of High COD. *Environmental Progress*. 22 (1): 57 – 65.
- [2] Can, O.T., M. Kobya, E. Demirbas, and M. Bayramoglu. 2006. Treatment of The Textile Wastewater by Combined Electrocoagulation. *Chemosphere*. 62: 181 – 187.
- [3] Carmona, M., M. Khemis, J. P. Leclerc, and F. Lapique. 2006. A Simple Model to Predict The Removal of Oil Suspensions from Water Using The Electrocoagulation Technique. *Chemical Engineering Science*. 61: 1237 – 1246.
- [4] Daneshvar, N., H. A. Sorkhabi, and A. Tizpar. 2003. Decolorization of Orange II by Electrocoagulation Method. *Separation and Purification Technology*. 31: 153 – 162.
- [5] Daneshvar, N., H. A. Sorkhabi, and M. B. Kasiri. 2004. Decolorization of Dye Solution Containing Acid Red 14 by Electrocoagulation with a Comparative Investigation of Different Electrode Connections. *Journal of Hazardous Materials*. B112: 55 – 62.
- [6] Daneshvar, N., A. Oladegaragoze, and N. Djafarzadeh. 2006. Decolorization of Basic Dye Solutions by Electrocoagulation: an Investigation of The Effect of Operational Parameters. *Journal of Hazardous Materials*. B129: 116 – 122.
- [7] Holt, P.K., G. W. Barton, M. Wark, and C. A. Mitchell. 2002. A Quantitative Comparison between Chemical Dosing and Electrocoagulation. *Colloids and Surfaces A : Physicochemical Engineering Aspects*. 211: 233 – 248.
- [8] Kobya, M., O. T. Can, and M. Bayramoglu. 2003. Treatment of Textile Wastewaters by Electrocoagulation using Iron and Aluminum Electrodes. *Journal of Hazardous Materials*. B100: 163 – 178.
- [9] Kobya, M., H. Hiz, E. Senturk, C. Aydinler, and E. Demirbas. 2006. Treatment of Potato Chips Manufacturing Wastewater by Electrocoagulation. *Desalination*. 190: 201 – 211.
- [10] Kumar, P. R., S. Chaudhar, K. Khilar, and C. Mahajan. 2004. Removal of Arsenic from Water by Electrocoagulation. *Chemosphere*. 55: 1245 – 1252.
- [11] Larue, O., E. Vorobiev, C. Vu, and B. Durand. 2003. Electrocoagulation and Coagulation by Iron of Latex Particles in Aqueous Suspensions. *Separation and Purification Technology*. 31: 177 – 192.
- [12] Lin, S. H., C. T. Shyu, and M. C. Sun. 1998. Saline Wastewater Treatment By Electrochemical Method. *Water Research*. 32 (4): 1059 – 1066.
- [13] Mollah, M. Y. A., R. Schennach, J. R. Parga, and D. L. Cocke. 2001. Electrocoagulation (EC) – Science and Applications. *Journal of Hazardous Materials*. B84: 29 – 41.
- [14] Mollah, M.Y.A., P. Morkovsky, J. A. G. Gomes, M. Kesmez, J. Parga, and D. L. Cocke. 2004. Fundamentals, Present and Future Perspectives of Electrocoagulation. *Journal of Hazardous Materials*. B114: 199 – 210.
- [15] X. Chen, G. Chen, and P.L. Yue. 2000. Separation of Pollutants from Restaurant Wastewater by Electrocoagulation. *Separation and Purification Technology*. 19: 65 – 76.

Integration of Constructability Principles into the Design Process

Woon, K. S.^{1*}, and Rosli, M. Z.²

¹ Faculty of Civil Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
E-mail: wkaisiong@yahoo.co.uk

² Faculty of Civil Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5532156, E-mail: rosli@fka.utm.my

Abstract

Constructability is an important element in building project's design phase, where designers' personnel play a prominent role to enhance it. Several researchers found that failure of design professional to consider constructability during the design phase can result many times of design reworks, contract changes, delay, increase of cost, and even legal entanglement and claims. This paper describes an on-going research that focus on establishment and integration of constructability principles into the current building's design process. Hence, the aims of this study are to recognise the local construction industry's current design process, in order to propose an integration of design phase's constructability principles to that existing design processes, and develop a constructability design review checklist for design checking. There are three distinct phases of this study: phase 1 involves literature review and preliminary interview; phase 2 consists of structured interviewing with design professional experts and models development; phase 3 comprises of constructability principles integration and checklist development. The data flow diagram (DFD) that used to model design processes is briefly described. Finally, the outcome of this study is establishment of existing design process models and integration of constructability principles, which performs in DFD form.

Keywords:

Constructability, Design Process, Data Flow Diagram

1. Introduction

Nowadays, constructability concept has been extensively being developed and applied in the USA, UK and later in Australia, where their studies have demonstrated that improved constructability has lead to significant savings in both cost and time required for completing construction projects [1, 2]. While in Malaysia construction industry, there is lack of available sources and reliable documentation that detail those constructability concepts and guide their application. Hence, Malaysian engineers have a disadvantage by not knowing what, when and how they should enhance the project constructability in design phase, compared with the engineers in more developed countries [3]. Consequently, on site frustrations are often occurred due to design decisions that lacked of knowledge regarding on how the object would be built. Therefore, it is important that the designers should incorporate constructability inputs and critiques as part of the design process in their design practice, where it is still the best time to influence project costs, decrease the likelihood of delays, contract change orders and legal

entanglement and claims [4, 5].

Besides, numbers of constructability improvement tools for project designs have been developed by some researchers [6, 7, 8], where the level of formality of those methods are varied. Nonetheless, constructability improvement tool in the form of checklist is considered to be comprehensive in term of the concepts covered [9]. Suitable constructability checklist for the local constructability industry is currently unavailable, therefore, as initially, it is essential to develop a constructability checklist that able to check a design work.

The objectives of this research are as follow:

- i. To determine the local construction industry's current design process flow.
- ii. To propose an integration of constructability principles into the existing design process flow.
- iii. To develop a constructability design review checklist.

2. Literature Review

Many research organizations and individual researchers have defined the constructability term in a number of ways.

* Corresponding Author. E-mail: wkaisiong@yahoo.co.uk

Construction Industry Research and Information Association [10] in UK defined constructability as the extent to which the design of a building facilitates ease of construction, subject to the overall requirements for the completed building. In U.S., Construction Industry Institute [11] defined it as the optimum use of construction knowledge and experience in the conceptual planning, engineering, procurement, and field operation phases to achieve the overall project objectives. While, among one of the individual researchers, Mendelsohn [5] defined constructability as the integration of construction expertise into the planning and design of a project so that the construction forces have the maximum opportunity to deliver the project in conformity with cost, quality, schedule and safety objectives of the project's stakeholders. Generally, all of these definitions stress the importance to achieve overall project objectives.

Guidelines are one of the most common constructability review methods can be found in construction industry. Such guidelines that commonly being used for constructability reviews in design phase are based on the constructability principles developed by CIRIA, CII and CIIA. These guidelines provided by those research organizations are aimed at providing designers with further details of main constructability principles that need to be considered during the design process [9]. Despite of the constructability guidelines are comprehensive in term of coverage, however, the applications of these constructability principles are too general and may make designers difficult to evaluate whether the project design has met the desired level of constructability. Therefore, further detail explanations and understanding are required the guidelines method is adopted.

Totally there are eighteen design phase constructability principles have been derived by Rosli [9]. They are:

- Carry out thorough investigation of the site;
- Design for minimum time below ground;
- Design for simple assembly;
- Encourage standardisation/repetition;
- Design for pre-fabrication, pre-assembly or modularisation;
- Analyse accessibility of the jobsite;
- Employ any visualisation tools such as 3D CAD to avoid physical interference;
- Investigate any unsuspected unrealistic of incompatible tolerance;
- Investigate the practical sequence of construction;
- Plan to avoid damage of work by subsequent operations;
- Consider storage requirement at the jobsite;
- Investigate the impact of design on safety during construction;
- Design to avoid return visit by trade;
- Design for skills available;
- Consider suitability of designed materials;
- Provide detail and clear design information;
- Design for early enclosure; and
- Consider adverse weather effect in selecting materials or construction methods.

These constructability principles are resulted from a collaboration of the design phase's constructability principles found by various researches and research organisations. In addition, these principles are more thorough with respect to ease of construction, and more focused and specified for the constructability improvement of design phase.

3. Research Methodology

Figure 1 shows the schematic of research methodology for this study. Generally, it consists of three distinct phases, named as phase 1, phase 2 and phase 3.

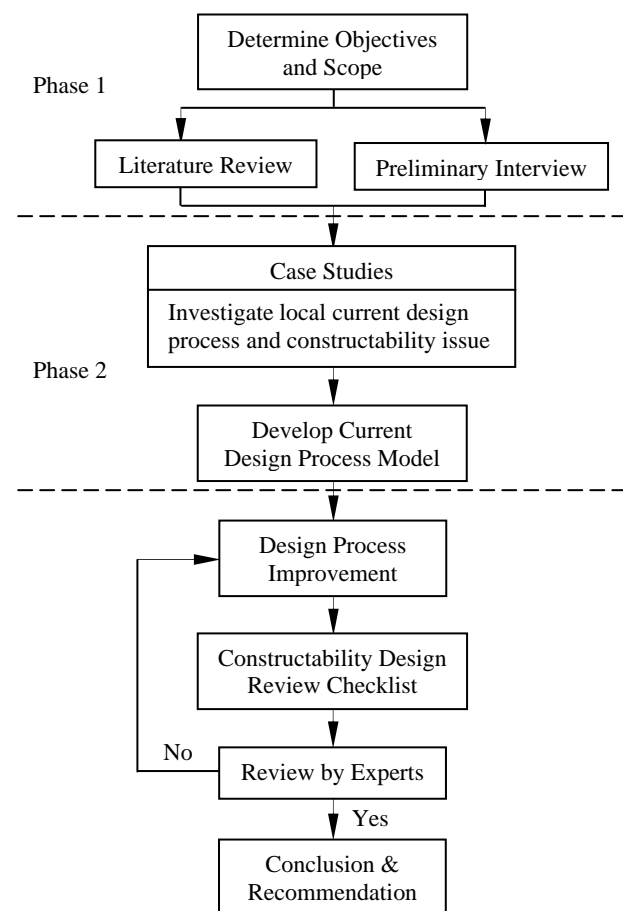


Figure 1. Schematic of research methodology

3.1 Phase 1

Extensive readings of previous studies and researches that related to constructability in project design phase had been carried out by the author in order to understand to the basics of the work. During the preliminary interview stage, the unstructured interview was adopted to obtain information and opinions from several designers and contractors with regard to design process flow, constructability and design-related problem. This interview was conducted by interviewing engineers who have minimum of four years experience in building design, and contractors who have minimum of four years experience in project construction.

3.2 Phase 2





According to Fischer and Tatum [12], interviews proved preferable to questionnaires as the knowledge acquisition method because they allowed direct interaction between the expert and the interviewer. Hence, for the case studies stage, structured interview will be used to collect detailed information about the current practice of the design process in building construction.

Three case studies are required to develop an existing design process model. Therefore, in this study, three consultant engineers who have at least eight years of experience in design process, and currently working and practicing their expertise in a civil and structure consultancy firm will be interviewed. These three designers will be interviewed with the objective of determining and identifying typical information related in the design process, which based on traditional procurement system of a building project. Besides, in order to gather more detailed information of design-related constructability problems in construction, the pre-appointed designers, one architect (has minimum of eight years experience in project design process), one contractor (has minimum of five years experience in building construction), and one clients will be interview for the opinion and suggestion. Those six interviewees will be regarded as the experts in this study. Table 1 summarized the basic selection criteria in determining the experts. Note taking and audiotape recording will be used to record interviewers' responses during the interviews due to its simplicity and reliability.

Table 1. Basic selection criteria in determining the experts

	No. of persons	Min. of experience	Types of firm
Civil engineer	3	8 years	C & S Consultant
Architect	1	8 years	Architect firm
Contractor	1	5 years	Building Contractor
Developer/ Owner	1	Not specified	Not specified

Table 2. Components of a data flow diagram for design (Adapted from Baldwin et al. [14])

Element	Software development description	Design process modelling description
Data or info. flow 	A connection between processes, etc., representing an input and/or an output	Design information flow
Process 	Individual functions that a system carries out. They transform an input into an output	Individual design task, e.g. calculation, drawing, specifying
Data store 	A collection of information that must be remembered for a period of time	Drawings, sketches, calculation files, reports, documents, specifications, computer files, etc.
Source/sink 	External entities with which the system communicates	Any external data source, e.g. client, local authority

After the interview stage, the collected information will be used to develop three current design process data flow diagrams which respectively represent to each case study. Then, these three project based design process models will be compared, combined and generalized to develop a general current design process model. The modelling technique will be used in this study is data flow diagram (DFD) and it is described at the following section.

3.3 Phase 3

After an existing design process model is developed, an integration of constructability principles will be proposed by inserting constructability design review processes into the existing design process model. Then, constructability reviews will be proposed at the certain stages of design completion, such as 0%, 30%, 60% or 90% of design completion. Due to limit of time, only a constructability design review checklist for building foundation design will be developed. This foundation design constructability review checklist acts as a tool for designers to check and asses their design works whether their foundations design considered and complied with those principles. However, the result for the constructability design checklist will not be presented in this paper.

Prior the final commencing of the works, all design process models, improved design process model, and foundation constructability design review checklist will be reviewed by the predetermined experts in order to check for its mistake and to ensure for its appropriateness.

4. Proposed Model

4.1 Data Flow Diagram

Data flow diagram (DFD) is used in this study to model the current design processes and information exchange between them. According to Ramli and Mesir [13], data flow diagramming is a means of representing a system at any level of detail with a graphic network of symbols showing data flows, data stores, data processes, and data sources/destination. The components of DFD and their interpretation in the design context are shown in Table

2. The processes represent design tasks. The data flow arrows indicate the flow or exchange of information and the data stores standards, specifications or design files. Data sources or sinks are used to represent external entities or organizations such as the architect, the client, the contractors or local authorities.

4.2 The Design Process Model

Generally, Figure 2, 3, 4 and 5 showed the initial establishment of building design processes which procured under traditional contracting system, with the proposed constructability review process. These figures illustrated the processes that partitioned in a top-down way. The top level of the model known as 'context diagram' or level 0, represents the building design process in the most general of terms. Then, the context diagram is subdivided and the processes further decomposed until the decomposition identified the tasks that produced specific information requirements. From that, it allows the higher regions of the model to be read to obtain an overview of the system, and, if more detail is required, the lower levels can be studied as and where necessary.

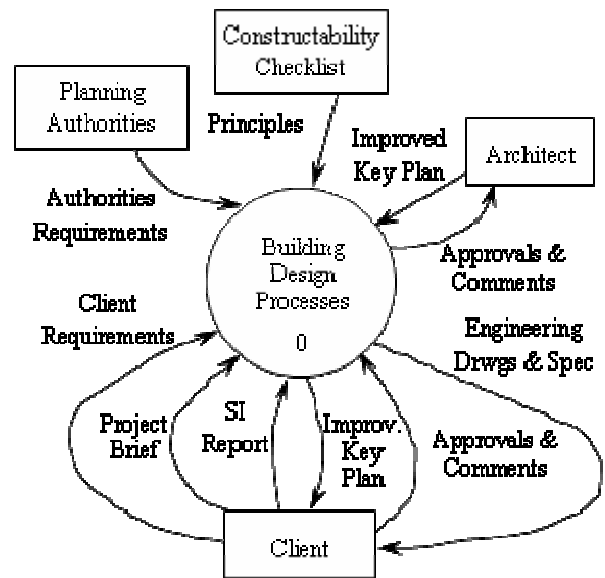


Figure 2. The Context Diagram

Level 1:

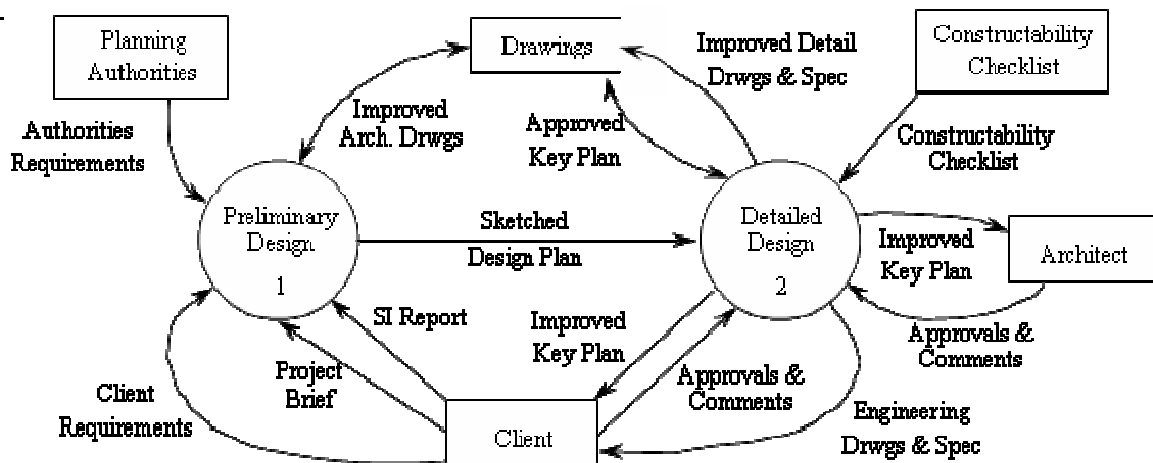


Figure 3. The Building Design Processes

Level 2:

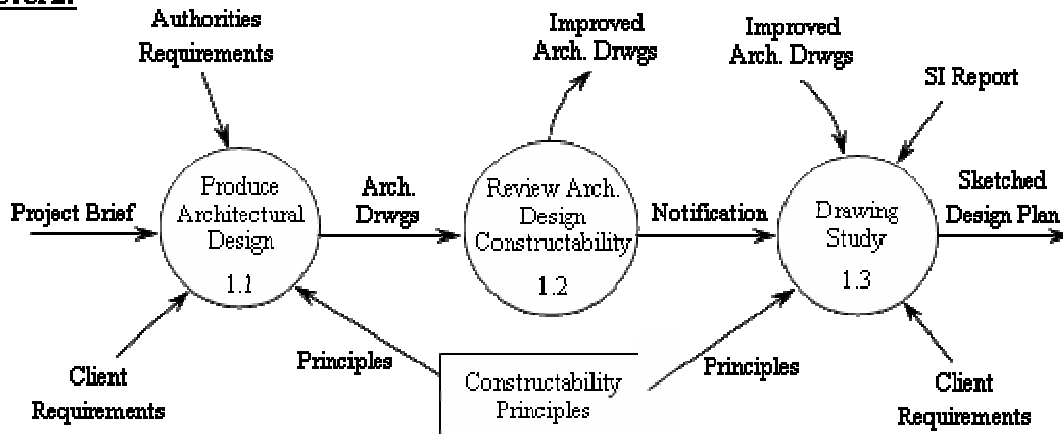


Figure 4. Process 1.0: Improved Preliminary Design

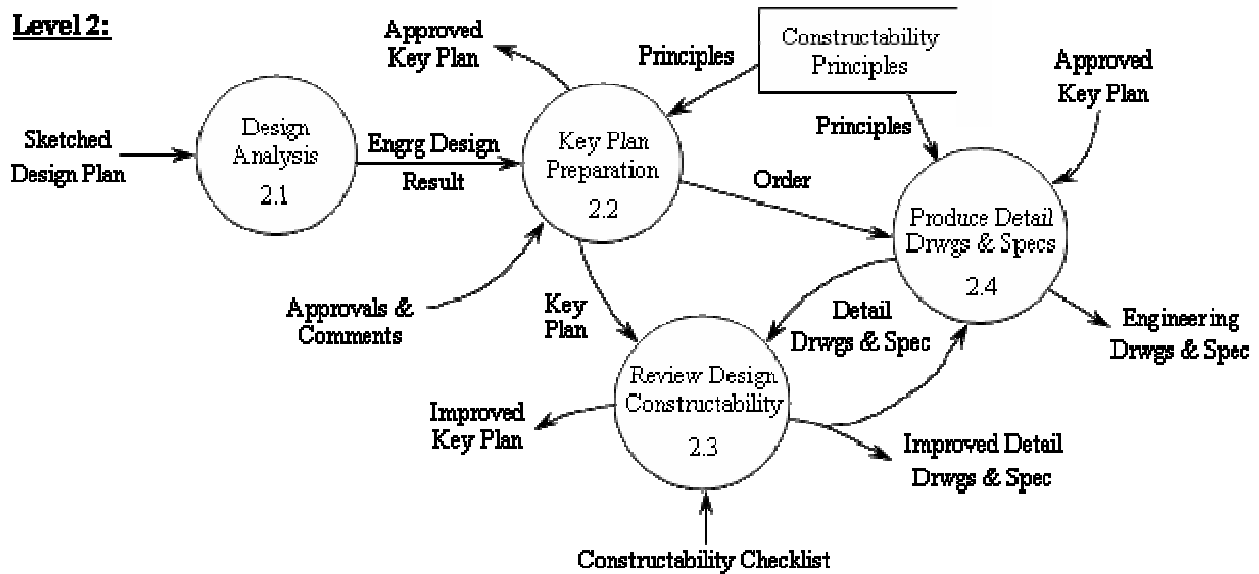


Figure 5. Process 2.0: Improved Detail Design

5. Conclusion

A large number of reworks during the building design of a project seems to be a most common practice in the local construction industry. This circumstance may be caused by many reasons, where lack in consideration of constructability during the project design is one of those reasons. Therefore, this paper demonstrated the initial development of current building's design process model that procured under traditional contracting system, with an integration of constructability principles into the design process flow. In order to encourage 'design correctly in first time', the constructability principles are inputted in the early stage of some processes. So that, each designer understands and knows the constructability principles that are required to be emphasized and implemented during their design. Nonetheless, several design reworks may be inevitable, but it is anticipated that the numbers of reworks can be minimized by implementing those design phase's constructability principles into the design practice.

References

- [1] Russell, J., Gugel, J., and Radke, M. W. 1992a. *Benefits of Constructability: Four Case Studies*. The Construction Industry Institute, Austin, Texas.
- [2] Jergeas, G., and Van der Put, J. 2001. Benefits of Constructability on Construction Projects. *Journal of Construction Engineering and Management*. 127(4): 281-290.
- [3] Nima, M. A., Abdul-Kadir, M. R., Jaafar, M. S., and Alghulami, R. G. 2001. Constructability Implementation: A Survey in the Malaysian Construction Industry. *Construction Management and Economics*. 19: 819-829.
- [4] Glavinich, T. E. 1995. Improving Constructability during Design Phase. *Journal of Architectural Engineering*. 1(2): 73-76.
- [5] Mendelsohn, R., 1997. The Constructability Review Process: A Constructor's Perspective. *Journal of Management in Engineering*. 13(3):17-19.
- [6] Navon, R., Shapira, A., and Sheehori, Y. 2000. Automated Rebar Constructability Diagnosis. *Journal of Construction Engineering and Management*. 126(5): 389-397.
- [7] Arditi, D., Elhassan, A., and Toklu, Y. C. 2002. Constructability Analysis in the Design Firm. *Journal of Construction Engineering and Management*. 128(2): 117-126.
- [8] Pulaski, M., and Horman, M. J. 2005. Organizing Constructability knowledge for Design. *Journal of Construction Engineering and Management*. 131(8): 911-919.
- [9] Rosli Mohamad Zin. 2004. *Constructability Assessment of Project at Design Phase*. PhD. Dissertation. Faculty of Civil Engineering, Universiti Teknologi Malaysia: Unpublished.
- [10] CIRIA. 1983. *Buildability: An Assessment*. CIRIA Special Publication 26. CIRIA, London.
- [11] CII. 1986. *Constructability – A Primer*. Publication 3-1. Austin, Texas: Construction Industry Institute.
- [12] Fisher, M., and Tatum, C. B. 1997. Characteristics of Design-Relevant Constructability Knowledge. *Journal of Construction Engineering and Management*. 123(3): 253-260.
- [13] Ramli, M. Z., and Mesir, B. 2004. *Civil Engineering IT for Civil Engineers*. Third Edition. UTM: ITUCE, Faculty of Civil Engineering.
- [14] Baldwin, A. N., Austin, S. A., Hassan, T. M., Thorpe, A. 1999. Modelling Information Flow during the Conceptual and Schematic Stages of Building Design. *Construction Management and Economics*. 17: 155-167.

The Effect of Joint Width between Blocks in Concrete Block Pavement

Rachmat Mudiyo^{1*}, Hasanan Md. Nor² and Tung Chai Ling³

²Faculty of Civil Engineering,
Universiti Teknologi Malaysia, 81310UTM Skudai, Johor, Malaysia
Tel: +60-7-5531500, Fax: +60-7-5566157, E-mail: hasanan@fka.utm.my

³Faculty of Civil Engineering,
Universiti Teknologi Malaysia, 81310UTM Skudai, Johor, Malaysia
Tel: +60-7-5531580, Fax: +60-7-5566157, E-mail: tcling611@yahoo.com

ABSTRACT

Jointing sand is the main component of concrete block pavement (CBP) that promotes load transfer between blocks. Often, when sands are available from various sources, it becomes a difficult task to judge the right source of sand. This paper presents the results of an experimental program, conducted to demonstrate a quick and simple technique, for selecting the right source of jointing sand for CBP. The optimum joint width between blocks is 3 mm. For joint widths less than the optimum, the jointing sand was unable to enter between blocks. A large amount of sand remained outside the joint sand heaps on the block surface.

Key words:

Concrete block pavement, width joint, jointing sand and sand grading.

1. Introduction

Concrete block pavement (CBP) consists of individual blocks of hand-sized units that are laid on a thin bed of sand, between edge restraints, overlaying a sub-base (Fig.1). The joint spaces are 2 to 7 mm and are filled with sand that is jointing sand. The presence of sand in the joints plays the biggest role in promoting load transfer between blocks. Frictional resistance is developed in the joints under load; this prevents the blocks from undergoing excessive relative displacements and transmits part of the load to adjacent blocks. The small shear displacements relative to each other facilitate the generation of horizontal forces between the blocks caused by dilatancy (i.e., the property to increase one's volume in the state of distortion) in the jointing sand. As a consequence, CBP is capable of achieving

substantial distribution of load among neighbouring paving units, due to increased frictional resistance. This paper presents the results of an experimental programme conducted to select the optimum joint width between blocks.

Jointing sand is the main component of CBP through which load is transferred to the larger area of lower layers by virtue of its shear and dilatancy property (Shackel 1985). Very few studies have been carried out concerning the suitability of sand for use in joints. Generally well-graded concreting sand is brushed and vibrated into the joints of CBP. In most cases, the sand used for bedding course is used in joint filling, which is not desirable. According to Knapton (1983), large joints require coarse sand and tight joints require fine sand for good performance of pavement.

* Corresponding Author. E-mail:
rachmatmudi@yahoo.com, H/P: 012-7312811

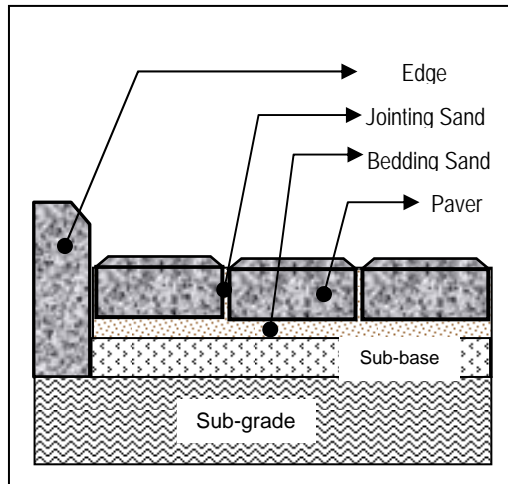


Figure 1. Structure of concrete block pavement

The shear strength of sand is basically made up of the structural resistance to displacement of the sand, because of the interlocking of sand particles, and the frictional resistance to translocation between the individual sand particles at their contact points. The jointing sand develops frictional forces under load, which resist the excessive relative displacement between blocks and transmit part of the load to adjacent blocks. Dilatancy means volumetric change accompanying deformation. This is a unique property of cohesionless soil. The term dilation is used herein to connote a thickening of a joint, i.e., an increase in the separation of the two joint blocks. Casagrande (1940) first described the influence of dilatancy on the drained friction angle of sands, while Rowe (1962) and Lee and Seed (1967) clarified the relative role of dilatancy, particle rearrangement, and particle crushing. Goodman and Dubois (1972) had incorporated the degree of dilatancy in analysis of jointed rocks. Shackel (1985) first mentioned the role of dilatancy in the jointing sand that facilitated the generation of horizontal forces between the concrete blocks and thereby increased the frictional resistance. During the process of shearing, the sand particles move over one another in the shear plane, showing an increase in thickness. A dilatant joint tested in a direct shear constrained to achieve constant normal deformation will have a higher angle of shearing resistance than one tested under constant normal stress (Goodman and Dubois 1972). In CBP, blocks are held tightly between rigid edge restraints. Adjacent blocks in the joint resist the thickening of the joint during dilation.

2. Experimental Programme

Materials

Sand: River sand from Kulai in Johor-Malaysia was used. The particle size distribution of bedding and jointing sand follow the grading requirement as tabulated in Table 1. Prior to use in each experiment, the sand was oven dried at 110°C for 24 h to maintain uniformity in test results. A maximum dry density of 1.73 gm/cc was obtained, corresponding to an optimum moisture content of 8.2 %. Two separate sand gradations were used for the bedding layer and in the block joints.

Table 1. Grading requirements for bedding sand and jointing sand

Sieve Size	Percent Passing For Bedding Sand	Percent Passing For Jointing Sand
3/8 in. (9.5 mm)	100	-
No. 4 (4.75 mm)	95 to 100	-
No. 8 (2.36 mm)	80 to 100	100
No.16 (1.18 mm)	50 to 85	90 – 100
No. 30 (0.600 mm)	25 to 60	60 – 90
No. 50 (0.300 mm)	10 to 30	30 – 60
No. 100 (0.150 mm)	5 to 15	15 – 30
No. 200 (0.075 mm)	0 – 10	5 – 10

BS 1377 Part I (1990) and TN 35: CCA (1996)

Paver: The concrete paving blocks conform to ASTM C 936, and 60 mm and 100 mm thickness, and rectangular in shape. Concrete blocks of 60 mm thickness and 110 x 220 mm of rectangular shape were used as the surface layer of the experiments. The mean compressive strength of the blocks was 32.30 MPa.

Test Set up

The horizontal force test was conducted using steel frame as edge restraint of 2.00 m wide and 2.00 m length. The test set up (shown in Fig. 2.) with construction varieties of CBP with laying pattern (stretcher bond, herringbone 90° and herringbone 45°) and joint width (3mm, 5 mm and 7mm). Loads were applied to the test pavement from side by using a hydraulic jacking system of 100 kN capacity clamped to the reaction steel frame.

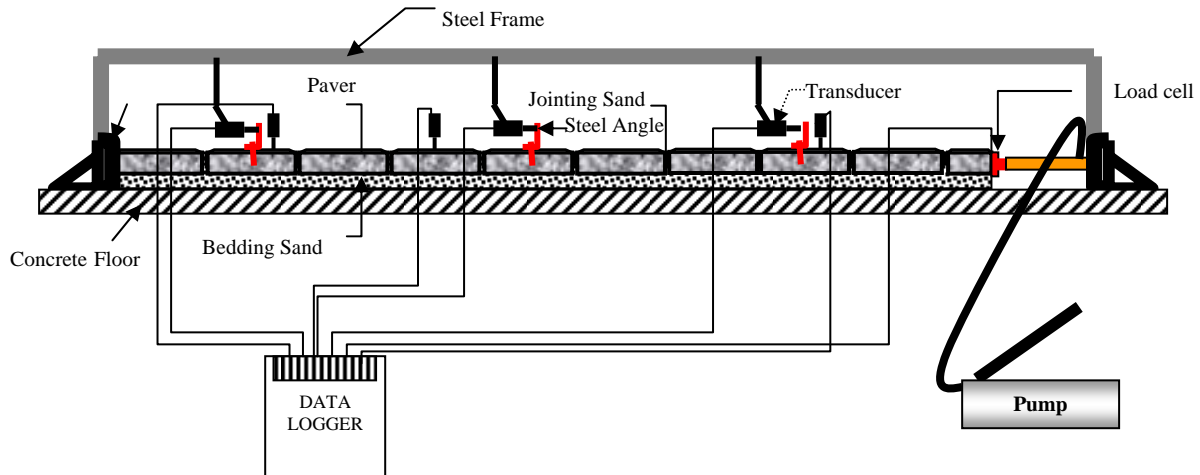


Figure 2. Horizontal Force Test Set up

The push-in test was conducted using steel frame in a laboratory-scale model assembled for this purpose (Fig.3). The test set up was a modified form of that used by Shackel (1980), where the pavers were laid and compacted within a steel frame in isolation from the bedding sand, sub-base course, and other elements of CBP. Here, instead of a frame, the tests were conducted in a box to incorporate the elements of CBP (i.e.

bedding course, jointing sand and paver). It consists of a rigid steel box of 1000 x 1000 mm square in plan and 200 mm depth, in which pavement test sections were constructed. The box was placed on a steel plate 10 mm thickness, beneath the reaction frame. Loads were applied to the test pavement through a rigid steel plate using a hydraulic jacking system of 100 kN capacity clamped to the reaction frame.

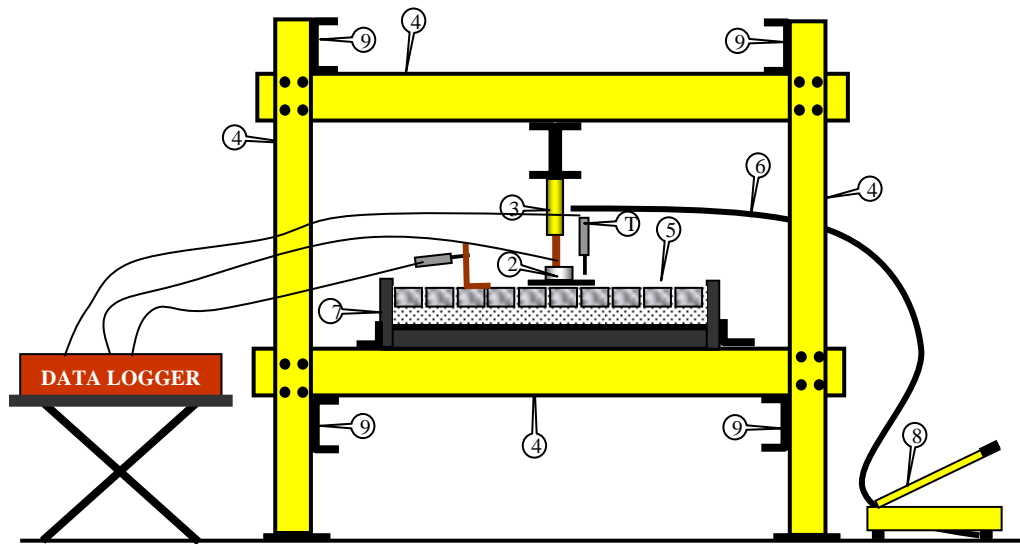


Figure 3. Push-in test setup

- | | | |
|-------------------|--------------------------|--------------------------|
| 1. Transducer | 4. Steel Frame C Profile | 7. Steel Frame Box 1mx1m |
| 2. Load Cell | 5. Bedding Sand | 8. Hydraulic Pump |
| 3. Hydraulic Jack | 6. Oil Pipe | 9. Cross Steel Frame |

3. Construction of Test Sections

The test sections of CBP were constructed within the box. A steel plate of 12 mm thick covered with sand was supported by steel frame (I beam profile). This thickness is reasonable for use in CBP to prevent immediate shear failure along the joints between blocks. At any two adjacent edges of the test pavement, side steel plates of required thickness were placed to control the desired width of joints in each test. The length and depth of side plates were 1000 and 200 mm, respectively. The depth was selected such that the side plates, when placed on the base, would reach the top of the block layer. Bedding sand of a particular gradation and thickness as per test requirements was uniformly screeded to a loose state. Pavers were then manually placed on the bedding sand in stretcher bond. Once the pavers were placed, they were compacted by a vibrating plate compactor of 250 N static weights vibrating at a frequency of 3,000 rpm. The compaction was continued until the top of each paving block was level with the adjacent blocks and to refusal of further settlement under vibration. The joints were then filled by brushing in the jointing sand. The joint filling operation was continued until all joints were completely filled with sand. Finally the top surface of pavement was cleaned of excess sand.

4. Test Procedures

A hydraulic jack fitted to the reaction frame applied a central load to the pavement through a rigid circular plate with a diameter of 300 mm. This diameter corresponds to the tire contact area of a single wheel, normally used in pavement analysis and design. A maximum load of 51 kN was applied to the pavement. The load of 51 kN corresponds to half the single axle legal limit. Deflections of the pavements were measured using four transducers to an accuracy of 0.01 mm corresponding to a load of 51 kN. The transducers were placed on two opposite sides of the plate at a distance of 100 mm from the center of the loading plate. The average value of four deflection readings was used for comparing experimental results.

The parameters, including joint width, thickness of bedding sand, and thickness of paver, were varied in the experimental program. For each variation of a parameter, the test was repeated three times to check the consistency of

readings. The average of the three readings is presented in the experimental results in graphical form. The range of the standard deviations (SD) of the readings for each parameter is presented in the respective figures. For each test, measurements of joint width were made at 20 randomly selected locations. The mean and standard deviation were calculated to assess the deviation from the design joint width. Design joint width as referred to herein be the desired width established in the experiment; however, the achieved joint widths always varied. The mean and standard deviation of joint widths description with and without sand before compacted are summarized in Table 2. While discussing experimental results, pavement deflections were compared referring to design joint widths.

Table 2. Width joint description with sand

Design Joint Width (mm)	Range of Joint Width (mm)	Standard Deviation of Joint Width (mm)
3	3.04	0.38
5	5.01	0.49
7	7.10	0.89

5. Results and Discussions

Sand was used both in the bedding and joints. The thickness of bedding sand was 30 mm, 50 mm and 70 mm. For each parameter, tests were carried out for 3 mm, 5 mm and 7 mm joint widths. Fig. 4 presents the deflections for pavements with and without application of jointing sand. For each joint width, the pavement without jointing sand deflected three times more than that of the pavement with jointing sand. This shows the importance of jointing sand. The concrete blocks in the pavement without jointing sand behave as individual units. Individual blocks do not transfer the applied load to adjacent blocks. Thus, the block layer has little load spreading capacity. The block layer obtains load spreading capacity if the individual blocks are interconnected. For this purpose, the joints between the blocks should be filled with sand.

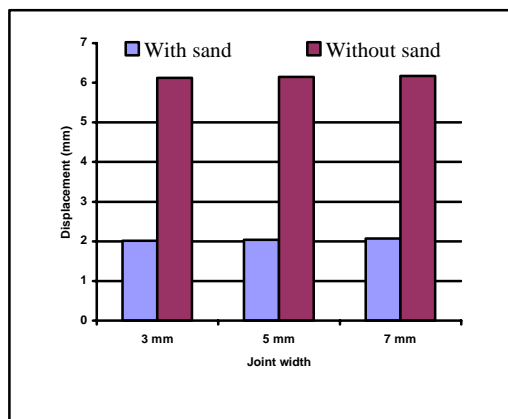


Figure 4. Pavement deflection with and without sand in joints

5.1. The Effect of Joint Width

The width of joints in block paving is more important than perhaps that has been realized in the past. A serious disadvantage of pavements laid in this way is that joints of less than 2 mm in width often contain little or no jointing sand. This would obviously reduce the contribution of individual blocks to the structural properties of the pavement. The individual blocks move in relation to one another which results in the spalling of the edges.

Sand filled joints are an integral part of concrete block pavement. They permit the block surface course to behave flexibly by allowing some articulation of individual blocks and they provide the structural interlock necessary for stresses to be distributed among adjacent blocks. Joints need to be sufficiently wide to allow this flexible behaviour, but not so wide as to permit excessive movement of the pavers. Joint widths should lay in the range of 2 to 7 mm with a preferred size of 3 mm. Those wider than 5 mm should not be accepted. 2 mm wider spacer ribs cast integrally on the vertical surfaces of the pavers ensure minimum joint width and assist in rapid placement of the blocks

The sand was used in bedding course with a 50 mm thickness for all of these experiments. Figure 5 shows the response of pavement for design joint widths of 2 mm, 3 mm, 5 mm, 7 mm and 9 mm with the same quality of sand. As the joint width decreases, the deflection of the pavement also decreases. The deflection of pavement decreases up to a certain point and then slightly increases with decrease in joint width, i.e., there is an optimal joint width.

The optimum joint widths for these experiments are 3 mm, respectively.

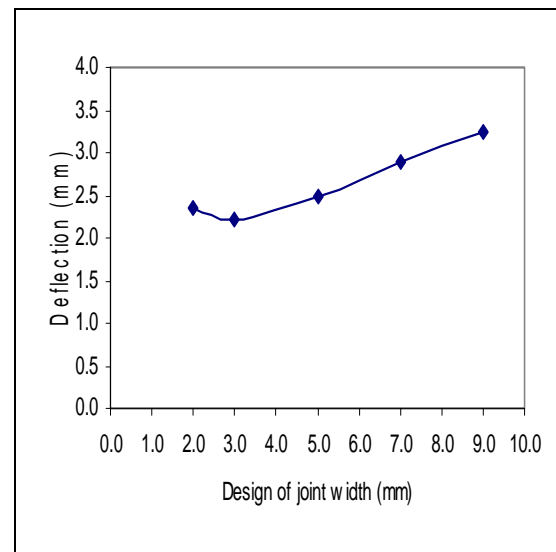


Figure 5. The response of pavement deflection for design joint widths

For joint widths less than the optimum, in a slight increase in deflections was observed. Some of the grains coarser than the joint width were unable to enter inside. This has been observed during filling sand in joints. A large amount of sand remained outside the joint showing sand heaps on the block surface. The coarse grains of sand choke the top surface of joints and prevent movement of other fine grains in to the joint. There might be loose pockets or honeycombing inside the joint. The joint stiffness decreases and in turn gives slightly higher deflections.

5.2. Filling of Jointing Sand

The compaction might not be fully effective for a higher thickness of bedding sand during vibration. The bedding sand rises through the joints to small heights and wedges in between the blocks. Figure 6 shows the rise of sand through the design joints width of 3 mm, 5 mm and 7 mm with varying thickness of bedding sand. The rise of sand increases with increase in thickness of bedding sand. The wedging of these sands absorbs the major part of applied vibration energy and transfers less to the bedding sand below. As a result, the bedding sand is not fully compacted for higher thickness (Shackel, 2003).

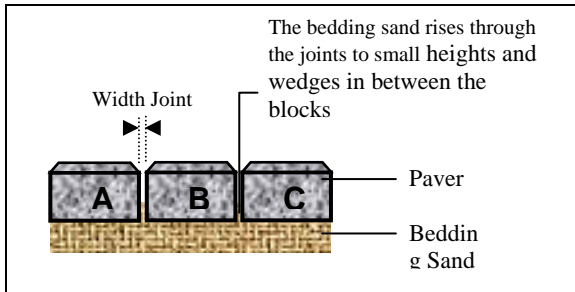


Figure 6. The bedding sand rises through the joints to small heights and wedges in between the blocks

Consequently, some compaction of bedding sand takes place under load and thus shows more deflection in the test pavements. The higher the bedding sand thickness, the more the deflection will be.

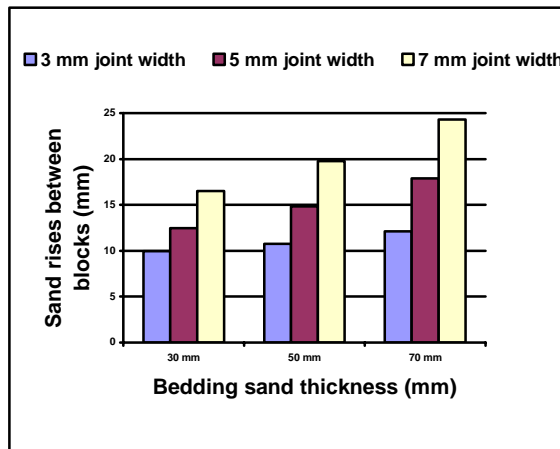


Figure 7. The comparison for bedding sand which rises in various widths joint and bedding sand thickness

The findings of this study are contradictory to those reported by Knapton and O'Grady (1983). They have found an increase in bedding sand thickness produced a proportionate increase in load-carrying capacity of pavement. As the pavement response is nearly for 50 mm thickness of bedding sand, it can be recommended for use in the field. But this depends on other factors, such as required level in sub-base tolerance and rise of bedding sand through the joints. Also, there should be a sufficient depth of bedding sand for deflection of pavements under load. Rise of bedding sand is essential to induce interlock.

The results show that the decrease in joint width increases the pavement performance and the concept of optimum joint width agrees well with that of a series of static load tests.

6. Conclusion

Based on the test results obtained in this investigation, the following conclusions can be drawn:

- In general, the coarser the sand the higher will be the dilatancy and angle of shearing resistance.
- Coarse-grained sand with a greater shearing resistance and a loose thickness of about 50 mm should be used in the bedding course of concrete block pavements. The maximum size of the jointing sand should be less than the joint width, and the sand should contain a small amount of fines (passing the 75 μ m sieve).
- The joints in between blocks should be properly filled with sand. The optimum joint width between blocks is 3 mm. For joint widths less than the optimum, the jointing sand was unable to enter inside between blocks. A large amount of sand remained outside the joint showing sand heaps on the block surface.

References:

- ASTM. (1972). Standard test methods for direct shear tests of soils under consolidated drained condition. *D 3080-72*, West Conshohocken, Pa.
- British Standards. (1973). Specification for aggregates from natural sources for concrete. *BS 882, 1201: Part 2*, London.
- Casagrande, A. (1940). Characteristics of cohesionless soils affecting stability of slopes and earth fills. *Contributions to soil mechanics 1925-1940*, Boston Society of Civil Engineers, Boston, 257-276.
- Goodman, R. E., and Dubois, J. (1972). Duplication of dilatancy in analysis of jointed rocks *J. Soil Mech. and Found. Div.*, ASCE, 98(4), 399-422.

- Knapton, J., and O'Grady, M. (1983). *Structural Behavior of Concrete Block Paving*. Journal Concrete Society: 17–18.
- Lee, K. L., and Seed, B. H. (1967). Drained strength characteristics of sands. *J. Soil Mech. and Found. Div.*, ASCE, 93(6), 117–141.
- Rowe, P. W. (1962). The stress dilatancy relations for state equilibrium of an assembly of particle in contact *Proc., Royal Soc. of London, Series A*, Vol. 269, London, 500–527.
- Shackel, B. (1980). The performance of interlocking block pavements under accelerated trafficking. *Proc., 1st Int. Conf. on Concrete Block Paving*, Newcastle-upon-Tyne, U.K., 113–120.
- Shackel, B. (1985). Evaluation, design and application of concrete block pavements. *Proc., 3rd Int. Conf. on Concrete Pavement Design and Rehabilitation*, Purdue University, West Lafayette, Ind., 113–125.
- Shackel, B., and Lim, D. O. O. (2003). *Mechanisms of Paver Interlock*. Proceeding 7th International Conference on Concrete Block Paving, South Africa.

An Overview of Construction Accidents in Malaysia

Abdul Rahim Abdul Hamid^{1*}, Muhd Zaimi Abd Majid¹, Bachan Singh¹

¹ Faculty of Civil Engineering

Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5531689, Fax: +60-7-5566157, E-mail: rahimfka@yahoo.com

Abstract

Malaysian construction industry is one of the main engines of growth of our domestic economy despite facing many challenges from the issues of low productivity and quality, poor image and safety, lack of ethics, data and information. Issues particularly related to the construction safety and health with regard to our construction industry is plentiful and seems to be never ending. This paper report the literature review of a postgraduate study which provide an overview of the current safety practices, construction accidents statistic, the impact of accidents and strategies to enhance construction safety in Malaysia. The result of the study shows that with regard to construction safety, over the years there have been increasing numbers of notification, compound and prosecution cases. The fatality rate from construction accidents are among the highest compared to the overall industry. Our government hope that the fatality rate of 25.8 per 100,000 workers in 2003 can be further reduced by 30% by the year 2010 or less than 20 per 100,000 workers by year 2020 (developed nation status) to be at par with the developed countries like Japan, France and the USA which fatality rate is below 20 per 100,000 workers. Accident contribute to the cost of construction, directly through increased compensation and insurance, and indirectly through decreased productivity, quality non-conformance and schedule overruns.

Keywords: Construction Industry, Accident, Construction Safety, Fatality Rate

1. Introduction

The construction industry is one of the main engines of growth of our domestic economy. Its cyclical growth is a barometric reflection of the economic climate, experiencing accelerated expansion when the economy is booming and sharp decline when the economy is on the downslide. Our construction industry has been facing many challenges from the issues of low productivity, low quality, fragmentation, poor image, economically volatile, bureaucratic delays, lack of ethics, shortage of skill manpower, and lack of data and information.

Generally, the production of construction products is a risky, complex and lengthy process. The total development of a construction project normally consists of several phases requiring a diverse range of specialized services. Cost, time, quality and safety are important characteristic of every project. For the construction industry in Malaysia, there has been greater emphasis on the first three aspects at the expense of safety. Many employers have not established comprehensive accident prevention policies but instead concentrate on maximizing profit. They do not emphasize on safety because they do not know how high the actual cost of an accident is until it occurs. Lack of adherence to safety

* Corresponding Author. E-mail: rahimfka@yahoo.com,
Tel: +60-7-5531689, Fax: +60-7-5566157

requirements has led to increased exposure of workmen and the general public to risky situation at construction sites resulting in a high chance of occurrence of accidents.

Issues particularly related to the construction safety and health with regard to our construction industry are often associated to high rate of accidents, low wages for high risk jobs, uncondusive work environment, competitive tendering, multi-layered subcontracting system, unskilled workers, high labor turnover, variable hazards, low priority for safety, harsh operating conditions, poor project and site management, low technology used, no centralize system to compile the scattered data and lack of relevant accidents data.

2. Methods

2.1. General Approach of the Research

This paper report a literature review for the postgraduate research on the causes of accidents and accident prevention strategies in the Malaysian construction industry. The research was started by identifying the broad research area which interest researcher and have the potential to be applicable to our construction industry. The research area chosen is in line with occupational health and safety (OHS) research categories adopted by Australian, United States, European Union, and World Health Organisation (WHO).

Initially, a preliminary literature review as well as interaction with safety experts and relevant agency such as National Institute of Occupational Safety and Health (NIOSH), Department of Occupational Safety and Health (DOSH), Social Security Organisation (SOCSO), safety academician, construction safety expert and other safety researchers has been conducted to enable the identification of key issues. Key words such as construction safety, safety, health, accidents, cause, construction sites, Malaysia, accident prevention, etc. related to the subject has been used to locate the available materials through the on-line databases system at the Universiti Teknologi Malaysia library. Later on a more extensive literature review has been performed through various database system and websites to acquire related articles, periodicals, books, reports, and serial, etc.

3. Results and Discussion

3.1. Industrial Accidents

The number of industrial accidents reported to the Social Security Organisation (SOCSO) has declined by 48

percent from 1993 to 2004 as shown in Table 1. SOCSO had received 133,293 reports on industrial accidents in 1993 and in 2004, the number had reduced to 69,132 cases [1]. In other words the industrial accidents for every 1,000 workers have declined from 21.2 in 1993 to 6.8 in 2004. However, if we compare this rate with that of the developed nation such as Sweden, Japan, Korea and European countries, there is still big task ahead to get the right figure. Our government aims to bring down the ratio of occupational mishaps to three (3) for every 1,000 workers before the country becomes fully developed [2].

On the other hand, the number of fatalities arising from industrial accidents for the same period has increased by 30 per cent from 653 in 1993 to 846 in 2004 [1]. The Fatality rate per 100,000 workers even though has marginally declined from 10.4 in 1993 to 8.3 in 2004 but the statistic show it has increased from the year 1995 until 2002. We need to bring down this rate to remain competitive or be at par with the developed nations which have low fatality rate such as Sweden (1.2), United Kingdom (1.3), Netherlands (1.9), Denmark (2.0), Finland (2.0), Australia(2.0), Japan (2.6) and Singapore (4.9) [3].

Table 1. Industrial Accident Rate

YEAR	REG. EMPLOYERS	REG. WORKERS	NO. INDUS. ACCIDENTS	ACC. RATE PER 1,000	NO. INDUS. FATALITY	FAT. RATE PER 100,000
2004	542,629	10,238,554	69,132	6.8	846	8.3
2003	507,853	9,996,848	73,858	7.4	822	8.2
2002	462,517	7,912,385	81,810	10.3	858	10.8
2001	443,904	8,769,321	85,292	9.7	958	10.9
2000	415,523	8,877,304	95,006	10.7	1004	11.3
1999	385,916	8,598,005	92,074	10.7	909	10.6
1998	358,543	8,428,589	85,338	10.1	1046	12.4
1997	338,794	8,252,680	86,589	10.5	1307	15.8
1996	305,500	7,613,635	106,508	14.0	1207	15.9
1995	274,017	7,412,191	114,134	15.4	952	12.8
1994	241,243	6,947,979	122,688	17.7	644	9.3
1993	212,932	6,285,194	133,293	21.2	653	10.4

Table 2. Construction Accidents Rate

Year	Registered Workers	Total Employment	Accident Reported	Permanent Disablement	Death	Acc Rate per 1,000	Fatality Rate per 100,000
2003	368,476	791,900	4,654	566	95	12.6	25.8
2002	349,084	782,100	5,015	652	88	14.4	25.2
2001	321,918	771,700	4,593	618	89	14.3	27.6
2000	278,127	828,000	4,873	642	159	17.5	57.2
1999	250,000	803,800	4,747	610	146	19.0	58.4

3.2. Construction Accidents

The number of construction accidents as shown in Figure 1 for the same period on the other hand has increased by 5.7 per cent from 4,207 cases in 1993 to 4,445 cases in 2004. The fatality rate as shown in Figure 2 has increased by 51 per cent from 51 cases in 1993 to 77 cases in 2004. Meanwhile, the cases for permanent disablement have increased by 46.1 per cent from 305 cases in 1993 to 566 cases in 2004. The fatality rate from construction accidents as shown in Table 2 are among the highest compared to the overall industry [4]. For the record, the construction accident rate per 1,000 workers are 19.0 (1999), 17.5 (2000), 14.3 (2001), 14.4 (2002) and 12.6 (2003). The fatality rate per 100,000 in construction are 58.4 (1999), 57.2 (2000), 27.6 (2001), 25.2 (2002) and 25.8 (2003). Our government hope that the fatality rate of 25.8 per 100,000 workers in 2003 can be further reduced by 30% by the year 2010 or less than 20 per 100,000 workers by year 2020 (developed nation status) to be at par with the developed countries like Japan, France and the USA which fatality rate is below 20 per 100,000 workers [5].

From the year 1999 to 2003 the employment in construction sector with regard to the overall industries is only at an average of 7.9% compare to agriculture (17.1%), manufacturing (25.8%) and services (48.8%) but the percentage of fatalities in the construction sector as shown in Figure 3 is always one of the highest at an average of 2.44% compare to agriculture (0.78%) and manufacturing (0.66%). International Labour Organisation (ILO) also highlighted the growing risks for workers worldwide especially in construction works where the rate of fatal accidents could be 10 or even 20 times higher than the average [6].

In 2003 the number of construction accidents reported to SOCSO is 4,654 cases out of 368,476 registered workers with SOCSO for Malaysian workers only. If we considered the estimated employment in this sector as reported to be around 791,900, then the accidents that went unreported could be higher. We have to bear in mind that 70 to 80 per cent of our construction workers are foreigner where some of them might have worked without or with expired work permits so it is generally believed that many cases of accidents and fatalities involving foreign workers are not

reported to the authorities. The actual figures are much higher if those not subscribing to SOCSO are taken into consideration. The published statistics are the tip of the iceberg. Similarly, research in United Kingdom (UK) by Health and Safety Executives (HSE) has shown that the accidents reporting level in construction industry is only at 55 per cent [7].

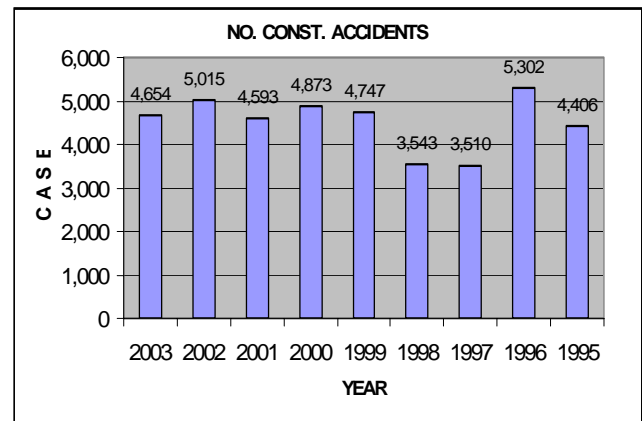


Figure 1. Number of Construction Accidents

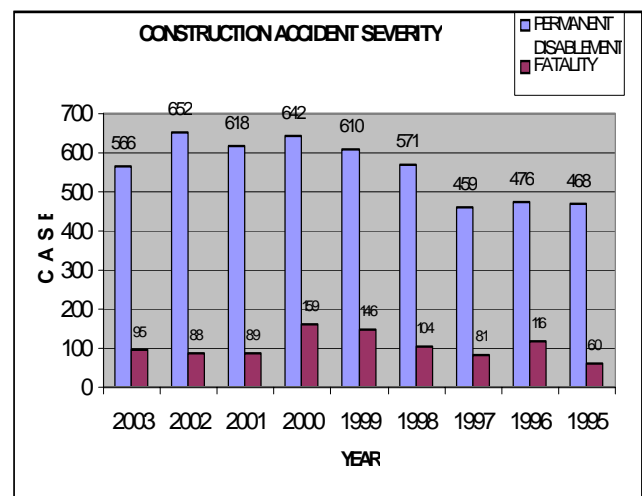


Figure 2. Construction Accident Severity

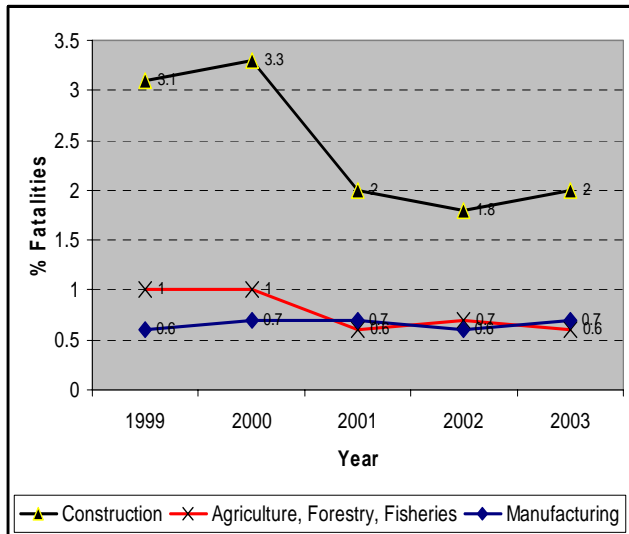


Figure 3. Fatality Rate by Sector

3.3. Cost of Accidents

The high accident rate in construction is nothing new but it is not something to be taken lightly. When accidents occur, it not only affects the injured person, but also the morale of other workers. Besides that, it also impairs the progress of the construction and costs money. Accidents not only give bad publicity but cause direct and indirect costs. One of the reasons is because time, cost and quality are always the main factors considered ahead of safety. Safety issues are always considered secondary and take a back seat in construction. Many employers don't realize the cost of an accident until it occurs.

To illustrate further let us consider the economic impact of accidents. Whenever accidents occur, it will generate a cost which can be categorized into a direct and indirect costs of accidents that can be large enough to affect the financial performance of any organization [8].

The direct costs include medical costs and other workers' compensation insurance benefits. For most construction companies, these direct costs of accidents are not fixed depending upon each company's own accident experience. Other direct costs of accidents are claims by

insured persons, which are settled either in or out of court, together with fines imposed by courts for breach of safety and health law. For construction companies with very poor safety records, the increased insurance premiums can be so expensive that they render a company sufficiently noncompetitive to put it out of business [8].

Indirect or hidden costs are the other and larger part of economic burden imposed by a poor accident record. There are many indirect costs of accidents such as reduced productivity, job schedules delays, added administrative time, damage to equipment and facilities as well as the less tangible, but very real, costs of human suffering, reduced morale, damage to the company image and so forth.

In term of cost of accident, according to SOCSO the number of recipients for compensation as shown in Table 3 has increased 36 percent from 182,763 people in 1995 to 247,790 people in 2003. In 2003 alone SOCSO has paid about RM 754 million, a staggering 161 percent increased from RM 289 million in 1995, as compensation for the industrial accidents. For the year 2004 the amount of compensation was RM 836 million payable to over 260,000 recipients [9]. Going by the Accident Iceberg Theory, the hidden or indirect costs of an accident is eight to 33 times more than that of its apparent or direct costs [10]. Therefore, just imagine the amount of hidden costs that we are spending yearly to finance these accidents and diseases.

The cost of industrial accidents was estimated to be between 14.5 and 18.1 billion pounds (in 95/96 prices) for the United Kingdom [11], 150 billion Dollars for the USA (in 1994) [12], between 1 to 3% of Gross National Product for the European Nations [13] and around 4% of world's Gross Domestic Product as mentioned by International Labour Organisation [14]. Whereas, in the United States (US) the total accident costs to the construction industry alone was estimated to be between US\$ 21 billion and US\$ 40 billion (in 1995 price) [12]. Research in the UK had found that the costs of accident in construction could be around 8.5% of the project cost and going by the same analogy it could cost EU nations around 75 billion Euros or almost €200 for every member of EU populations [15].

Table 3. Yearly SOCSO Compensation

Year/ Amount	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Cases	114,134	107,635	86,589	85,338	95,006	92,074	85,869	84,810	73,858	69,132
Compensation RM - Million	289	316	384	445	497	609	638	722	754	836
Yearly % change		9	22	16	12	22	5	13	5	11

Table 4. Statistics of Non-compliance

Year	2000	2001	2002	2003
Stop work order and notice of improvement	1023	4797	5746	5344
Compound	5	121	927	1117
Charges in court	24	58	76	121
Industrial accidents investigated	1171	1123	1231	1214
Complaints investigated	289	347	395	418

3.4. Level of Compliance

In another aspect of level of compliance, inspection conducted by the Department of Occupational Safety and Health (DOSH) from year 2000 to 2003 reveal that the achievement of satisfactory level of the assessment on the occupational safety and health at construction sites throughout the country were only between 50% and 80% [16]. These figures could indicate higher compliance by the contractors to safety and health procedures but that does not necessarily mean the overall safety performance at our construction sites has reach the satisfactory level. The statistic has proven otherwise. Some of site that were in the satisfactory level were just at the passing marks. Inspections were not comprehensive as they were conducted at random sites four times a year and the focus was more on the five worst negative elements namely unfenced edges, working at height, scaffolding, holes and platform.

The assessment criteria was based mainly on the compliance to certain aspects of the Building Operation and Work of Engineering Construction 1986 (BOWEC 1986) Regulations and the Occupational Safety and Health Acts 1994 (OSHA 1994) and not on the overall safety performance of construction sites. In addition to that, report from DOSH had shown that there were increasing numbers of notification, compound and prosecution cases for the year 1999 to 2003 as shown in Table 4. As a result of December 30th 2005 highly publicized accident at Plaza Damas involving a young corporate figure, Construction Industry Development Board (CIDB) was reported to reveal that there are 700 cases of negligence at project sites, which include 150 high-profile incidence since 2003 [17].

3.5. Accident Prevention Strategies

Management can prevent or mitigate accidents by identifying and implementing appropriate control barriers, which help to prevent errors or failures that could result in an accident. Accidents occur when one or more barriers in a work system, including procedures, standards and requirements intended to control the actions of workers, fail to perform as intended.

Suraji et al (2001) pointed out that from the analysis of 500 accidents records provided by the UK HSE, the most frequent category of proximal cause is inappropriate construction operation, occurring in 88% of all accidents. The causes are mainly attributed to site management during

construction [18].

Abdelhamid and Everett (2001) also noted that the ability of the designer to influence what happens at the construction site typically ends when the construction contract is awarded. Even if such influence does exist, a designer cannot be held accountable for an accident caused by a worker's deliberate unsafe action. In case cases, the onus is on on-site management to find ways to correct and discourage these actions and prevent their reoccurrence [19].

Heinrich et al (1980) suggested various short and long term accident prevention strategies. Short term strategies should focus on accidents and incidents, unsafe acts and conditions, company and safety rules, procedures and equipments. Long term strategies on the other hand will focus on the company policy and climate, safety program, management style, motivation, employee needs, morale, training, etc [20].

The situation in the Malaysian construction industry underpins the need to increase research activities and strengthen safety and health activities as well as to identify strategies to reduce the fatalities, injuries and ill health. For that reason Construction Industry Development Board has come out with a Master Plan for Occupational Safety and Health in Construction Industry 2005-2010 which spell out the strategies to improve the standard of safety and health in our construction industry. The strategies will focus on six core area which are enforcement and legislation, training and education, promotions, incentives and disincentives, standards, research and development and technology [5].

4. Conclusion

The statistics of accidents occurred in the construction industry indicate that the accident rate in our construction industry is still high. If we compare those rate with that of the developed nations there are still big task ahead to get to the right figure. The statistics discussed is to give us a clear picture that Malaysian construction industry is one of the critical sectors that need a huge and fast overhaul from the current site safety practices. Accidents could result in not only direct physical injury to persons or damage to property, but also short and long term effects or incidents due to other exposures on sites that affect the workers' health and physical well-being. A civilized society reacts to disaster before it happens. In order to prevent an accident, effective strategies must be formulated and implemented. For

example the root causes of construction site accidents must be investigated for an accident to be predictable. The knowledge of accident causation can be utilised in formulating a more conducive working conditions and environments at construction sites and hopefully the number of construction accidents could be minimised.

References

- [1] PERKESO. 12 May 2001. Statistik Kemalangan Industri URL <http://www.perkeso.gov.my/statistik.htm>.
- [2] DOSH. 19 Oct. 2000. Teks Ucapan YAB Dato' Seri Abdullah Ahmad Badawi Kempen Bulan Keselamatan dan Kesihatan Pekerjaan 2000. URL <http://www.dosh.gov.my/maklumat/ucapan/10200019tp mukempen.html>.
- [3] Asiah Jaafar, 2005. OSH As A Strategy Towards Excellent Performance in Public Sector. In *Proceedings of 8th NIOSH Conference And Exhibition On Occupational Safety And Health (COSH 2005)*, pp 96. PWTC, Kuala Lumpur: National Institute of Occupational Safety and Health.
- [4] NSTP. 9 April 2001. No Induction Course Attendance, No Construction Workers, Developers, Contractors Have Until July 1 to Comply URL <http://www.aboutsafety.com/article.cfm?id=683>.
- [5] Construction Industry Development Board Malaysia (CIDB). 2005. *Master Plan for Occupational Safety and Health in Construction Industry 2005-2010*. CIDB: Kuala Lumpur.
- [6] ILO. 22 April 1996. Press Release: ILO Highlights Growing Risks to Worker Health and Safety. URL <http://www.ilo.org/public/english/bureau/inf/pr/96-13.htm>.
- [7] HSE. 29 April 2001. Levels and Trends in Workplace Injury: Reported Injuries and the Labour Force Survey 1997/1998 URL <http://www.hse.gov.uk/laborsurvey>.
- [8] Levitt, R.E. and Samelson, N.M. 1993. *Construction Safety Management*. Second Edition. John Wiley & Sons, Inc.:USA
- [9] MOHR. 18 October 2004. Datuk Wira Dr Fong Chan Onn - Minister Of Human Resources, "Speech in Majlis Penyampaian Faedah PERKESO Tampin, Negeri Sembilan", URL <http://www.mohr.gov.my>.
- [10] MOHR. 14 July 2003. YB Datuk Dr Fong Chan Onn Minister Of Human Resources, "Speech in National Conference & Exhibition On Occupational Safety & Health (COSH 2003) Sunway Pyramid Convention Centre", URL <http://www.mohr.gov.my>
- [11] HSE. Press Release E207:99-25 Oct. 1999 Work Accident and Ill Health Cost Society Billions – New HSE Report Estimates URL <http://www.hse.gov.uk/press/e99207.htm>
- [12] Everett, J.G. and Frank, P.B. 1996. Costs of Accidents and Injuries to the Construction Industry. *Journal of Construction Engineering and Management, ASCE*, Vol 122, No. 2, Paper No. 10616.
- [13] European Agency for Safety and Health at Work (EU ASHW) (2002), Inventory of Socioeconomic Costs of Work Accidents, Luxembourg, pp5.
- [14] International Labour Organization (ILO). 2005. Prevention: A Global Strategy – Promoting Safety and Health at Work, The ILO Report for World Day for Safety and Health at Work Geneva, 2005, pp3.
- [15] European Agency for Safety and Health at Work (EU ASHW) (2004), European Week for Safety and Health at Work 2004: Why Construction - Financial Costs URL http://osha.eu.int/ew2004/static/whyconstruction/index_en.htm
- [16] NSTP. Building Sector to Achieve 'Zero Accident' Rate by 2003. 19 April 2001. URL <http://www.aboutsafety.com/article.cfm?id=726>.
- [17] STAR. December 2006. What Went Wrong at 700 Sites, Asks Fong. URL <http://www.mystar.com.my/article.safety.htm>
- [18] Suraji, A., Duff, A.R., and Peckitt, S.J., 2001. Development of Causal Model of Construction Accident Causation, *Journal of Construction Engineering and Management*. ASCE Vol 127, No. 4, July/August, 2001, pp337-344.
- [19] Abdelhamid, T.S. and Everett, J.G. 2000. Identifying of Root Causes of Construction Accident, *Journal of Construction Engineering and Management, ASCE*. January/February 2000, pp.52 – 60.
- [20] Heinrich, H.W., Petersen, D. and Roos, N. 1980. *Industrial Accident Prevention: A Safety Management Approach*, Fifth Edition, McGraw Hill Book Co: New York.

Predicting the Performance of Design-Bid-Build Projects: A Neural-Network Based Approach

Caren C. L. Tan¹, Rosli M. Z.²,

¹ Faculty of Civil Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
E-mail: tan_caren@hotmail.com

² Faculty of Civil Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5532156, E-mail: rosli@fka.utm.my

Abstract

Several studies had shown that many project managers are facing difficulties in predicting the performance of Design-bid-build (DBB) projects. This is due to the fact that there are many factors that affect DBB project success. This paper describes an ongoing research that is carried out to identify these factors. In addition, a proposed model to predict the performance of DBB project based on time is also described. Through literature research, a total of forty-four factors that affect DBB project success had been established. The degree of importance for these factors had been determined through questionnaire survey. Eight out of forty-four factors that affecting project performance were found to be the most important factors from the viewpoint of project managers and contractors in the Malaysia construction industry. The outcome of the survey formed a basis for the model development. Artificial Neural Network (ANN) will be adopted to map the relationship between the input and output variables. It is anticipated that the performance prediction model can be a useful tool to various parties in the construction industry.

Keywords: Artificial Neural Network; Project performance; Design-Bid-Build.

1. Introduction

Construction projects are intricate, time-consuming undertakings. The total development of a project normally consists of several phases requiring a diverse range of specialized services. Traditionally, field construction is not begun until the architect-engineer has completed and finalized the design. This sequence is still predominant in the industry and is referred to as the design-bid-build procedure. It is possible to reduce the total design-bid-build time by starting the construction before completing the design of the entire project. Measurements of performance provide management with invaluable feedback to guide daily decision making and they become more competent. On-time completion means that the job finished as it was scheduled. However, time and budget measurements frequently come too late to guide daily management decision making.

Studies had shown that project managers always encounter difficulties to predict the performance of DBB project. They need the skills to evaluate the factors that affect DBB project success. Under this circumstances, there is an urgent need to study, analyze and investigate into the prevailing conditions with a view to identify the factors and formulate solutions and, if required, to carry out extensive researches to upgrade and improve the project performance evaluation.

The objectives of this research are stated as follow:

- i. To identify factors that affecting the project performance.
- ii. To determine the degree of importance for each of the respective factors.
- iii. To develop a model based on Artificial Neural Networks (ANN) to predict the performance of design-bid-build (DBB) projects from the time aspect.

2. Literature Review

Generally, performance of a construction project can be evaluated by time, cost and quality (Konchar and Sanvido, 1998). For the purpose of this research, the performance of DBB projects will be measured by time aspect. Time was chosen as the measurement of performance after preliminary interview with several project managers being carried out. Time is an important factor as it is most concerned by the project manager and client in construction site. Time performance will be evaluated based on 3 aspects, which are behind time, on time and ahead time. If the project is ahead time, that means the progress of project is ahead the planned schedule, so the project is good and satisfying.

Several empirical studies relevant to the identification of factors influencing project performance were reviewed. Researches had identified more than forty factors

contributing to project success. A total numbers of forty-four factors affecting the project performance are listed in Table 1 (Albert et al., 2004). The factors affecting project success are categorized into attributes relating to the

project characteristic, project procedures, project management actions, project participants, and external environment.

Table 1: Factors affecting project performance

Project Aspect	Factors Related
Project Characteristic	<ol style="list-style-type: none"> 1. Type of project 2. Nature of project 3. Number of floors of the project 4. Complexity of project 5. Size of project
Project Procedures	<ol style="list-style-type: none"> 1. Procurement method 2. Tendering method
Project Management Actions	<ol style="list-style-type: none"> 1. Communication system 2. Control mechanism 3. Feedback capabilities 4. Planning effort 5. Developing an appropriate organization structure 6. Implementing an effective safety program 7. Implementing an effective quality assurance program 8. Control of subcontractors' work 9. Overall managerial actions
Project Participants	<ol style="list-style-type: none"> 1. Client's experience means whether he is a sophisticated or specialized client. 2. Nature of client means whether he is privately or publicly funded. 3. Size of client's organization. 4. Client's emphasis on low construction cost. 5. Client's emphasis on high quality of construction. 6. Client's emphasis on quick construction. 7. Client's ability to brief. 8. Client's ability to make decision. 9. Client's ability to define roles. 10. Client's contribution to design. 11. Client's contribution to construction. 12. Project team leaders' experience. 13. Technical skill of the project team leaders. 14. Planning skill of the project team leaders. 15. Organizing skill of the project team leaders. 16. Coordinating skill of the project team leaders. 17. Motivating skill of the project team leaders. 18. Project team leaders' commitment to meet cost, time and quality. 19. Project team leaders' early and continued involvement in the project. 20. Project team leaders' adaptability to changes in the project plan. 21. Project team leaders' working relationship with others. 22. Support and provision of resources from project team leaders' parent company.
External environment	<ol style="list-style-type: none"> 1. Economic environment 2. Social environment 3. Political environment 4. Physical environment 5. Industrial relations environment 6. Technology advanced

Adapted from: Albert et al., 2004

3. Research Methodology

Forty-four potential factors affecting project performance were identified from the review of past works. A set of self-administered questionnaires was developed to facilitate systematic data collection. The questionnaires can be divided into two parts, which were Questionnaire-Part1 and Questionnaire-Part2.

The Questionnaire-Part1 was designed to determine the degree of importance for each of the factor over the forty-four factors. This has enabled the most important factors that brought most impacts in influencing the construction performance to be determined. Sixty sets of questionnaires had been distributed to developers and contractors who had participated in DBB projects before. In order to obtain the data, all the respondents had a choice of being interviewed or self-administer the questionnaires, and then sending them back to the researchers. Respondents were requested to rate the degree of importance for each of the potential factors that will affect the project performance according to a five-point Likert scale. The scores were given from 1 to 5, as below:

Strongly Agree	5
Agree	4
Neutral	3
Disagree	2
Strongly Disagree	1

If two answers were selected or no answer was indicated, it will be regarded as invalid. Invalid answers were treated as missing cases and excluded from analysis. All data obtained from the respondents had been recorded. In the analysis process, the forty-four potential factors had been evaluated based on Average Index in order to identify the most important factors from the aspect of project managers. The average index is calculated as follow:

$$\text{Average Index} = \frac{\sum a_i x_i}{\sum x_i} \quad (1)$$

where : a_i = Constant expressing the weight given to i
 x_i = Variable expressing the frequency of response for $i = 1,2,3,4,5$

There are five categories of scale based on Average Index (Abd. Majid, 1997), which are:

1.00-1.49	Not Important
1.50-2.49	Slightly Important
2.50-3.49	Moderately Important
3.50-4.49	Very Important
4.50-5.00	Most Important

The factors that are fallen in 4.50-5.00 categories had been considered as the most important factors. Then the most important factors will be used for developing the Questionnaire-Part 2.

In the second survey, respondents will be requested to rate their own project based on the input to the identified most important factors according to behind time, on time or ahead time. If the project is ahead time, which means the progress of project is ahead the planned schedule, the project progress is considered good and satisfying. The data collected will be used to develop project performance

prediction model based on Artificial Neural Network (ANN) technique. After that, the model will be tested and validated based on the error generated in each model (Mean Square Error). The flow chart of research methodology for this study is shown in Figure 1.

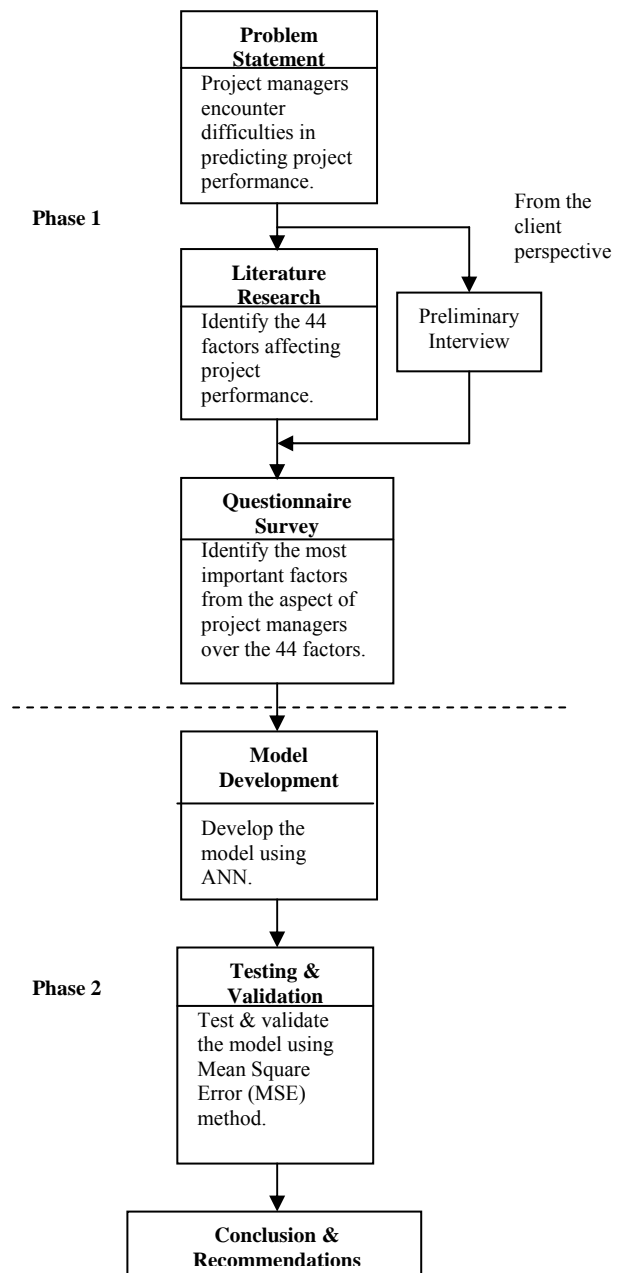


Figure 1: Flow chart of research methodology

4. Proposed Model

a. Model Development

ANN has a robust learning capability for it can reasonably validate whether all the variables are indeed the key determinants that affect project performance. It can produce fairly accurate predictions also, even when the information provided is not sufficient. Not only that, it can automatically approximate whatever functional form best characterizes the data. Backpropagation neural network architecture will be

chosen for constructing the ANN model. Through an iterative process, the neural network is able to learn from a subset of the data collected (as knowledge domain) represented by case examples and draw its own pattern of relationships. Using this approach, the predictive model will be ensued.

b. Data Collection

Proceed from data collection of second questionnaire survey; the input for the ANN model will be performance factor while the output will be performance measure based on time. The developed ANN model will be consisted of an input layers (performance factor), one or more hidden layers and an output layer (performance measure) as shown in Figure 2. The input layer represented the data to the network. Hidden layers acted as layers of abstraction and pull features from inputs. The number of nodes in the hidden layer was decided by trial and error. The output layer had three nodes, which were ahead time, on time and behind time.

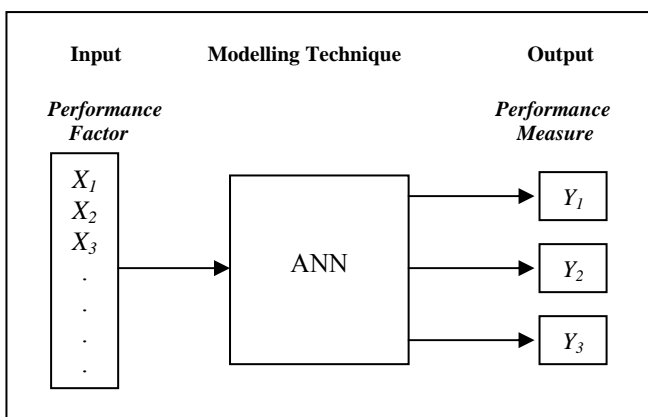


Figure 2: Proposed Model

c. Training and Testing

ANN estimates the time performance by learning from numerical examples and representing complex nonlinear relationships of these examples. Training will be used to train the application while testing will be used to measure the performance of a trained application. To develop the model, 85% of the data will be used for training and validation process while 15% of the data will be used for testing.

During training, neural techniques need to have some way of evaluating their own performance. Since they are learning to associate the inputs from the training with their outputs, evaluating the performance of the application on the training data may not produce the best results from the system. This is because if a network is left to train for too long, it will overtrain. This means that it will lose its ability to generalize. So that the neural computing technique can monitor its performance in a more sensible fashion, another part of the data is set aside as a validation set.

d. Validation

ANN models have the self-learning ability, by adjusting their parameters to reduce the error of estimation. Once the learning process had finished and the weights of the neural network had been calculated, it is important to check the quality of the resulting model. For example, in the case of supervised learning, a measure of the quality can be given in

terms of the errors between the desired and the computed output values for the training data. The standard error measurement method that will be used in this research is Root Mean Square Error (RMSE) Method, expressed by Equation, which can be defined as:

$$\sqrt{\sum_{p=1}^r \|b_p - B_p\|^2 / r} \quad (2)$$

Where: B_p = Actual duration to accomplish the project;
 b_p = Predicted duration to accomplish the project;
 r = Total number of cases.

e. Tool Used In Developing ANN Model

In this research, the prediction model will be developed using SPSS: Neural Connection software version 1.0. Neural Connection software offers a unique technology that recognizes complex patterns and trends in data, much like the human brain. To develop a predicting model using Neural Connection, the procedures can be broken down into a number of steps:

- i. **Pre-project:** Setting out project goals, establishing the desired results and system requirements.
- ii. **Data collection:** Ensuring that the correct data are gathered.
- iii. **Data preparation:** Cleaning the data and ensuring that they are in the appropriate format.
- iv. **Design:** Choosing the best neural approach.
- v. **Training and testing:** Building the application.
- vi. **Experimentation:** Tailoring the application to improve results.
- vii. **Implementation:** Producing the results.

5. Results

Thirty-two sets of Questionnaire-Part1 were received out of sixty sets of distributed questionnaires. Questionnaire analysis had been carried out. All data obtained from the respondents were recorded accordingly. After that, the data were tabulated using Microsoft Excel based on the Likert scale. Then the analyzing process was carried out. The forty-four potential factors, which listed in Table 1, had been evaluated using Average Index. Table 2 showed the most important factors resulted from the analysis, which the average index for these eight factors were fallen in the category of 4.50 – 5.00.

Table 2: The most important factors resulted from average index

No.	Factors	Average Index
1.	Complexity of project	4.59
2.	Control of subcontractors' work	4.75
3.	Client's emphasis on quick construction	4.69
4.	Project team leaders' experience	4.91
5.	Technical skill of the project team leaders	4.94
6.	Planning skill of the project team leaders	4.91
7.	Coordinating skill of the project team leaders	4.91
8.	Project team leaders' adaptability to changes in the project plan	4.75

5.1 Input and Output Variables

The outcome of the Questionnaire Survey – Part 1 had formed a basis for the model development. Table 3 provides an overview of the input and output variables in the ANN Prediction Model. The ANN prediction model has an input layer, one or more hidden layer(s) and an output layer. In this research, the input layer consists of eight inputs (see Table 3),

which are the performance factors while the output layer consists of three outputs, which are the performance measure. The hidden layer(s) act as the processing method.

5.2 The ANN Model

Back-propagation (BP) network is one of the simplest and most practicable network being used in performing higher level human tasks such as diagnosis, classification and forecasting. In view of the ability of BP network to perform various tasks coupled with the fact that it is one of the easiest networks to understand, Multi-Layer Perceptron (MLP) will be adopted as the neural computational technique. Figure 3 illustrates the architecture of MLP project performance prediction model.

6. Conclusions

The development of a model to predict the performance of DBB project based on time has been the focus of the research described in this paper. It is based on the most important factors affecting the project performance identified through questionnaire survey. The proposed Artificial Neural Network (ANN) model consists of an input layer with eight input variables, a hidden layer, and an output layer with three output nodes.

Table 3: Variables for ANN Prediction Model

Var ref	Explanatory variables	Definition
INPUT (Performance Factor)		
X ₁	Complexity of project	Scale 1 – 5; 1 = Not Complex; 5= Highly Complex
X ₂	Control of subcontractors' work	Scale 1 – 5; 1 = Poor; 5 = Excellent
X ₃	Client's emphasis on quick construction	Scale 1 – 5; 1 = None; 5 = Very High
X ₄	Project team leaders' experience	Scale 1 – 5; 1 = No Experience; 5=Highly Experience
X ₅	Technical skill of the project team leaders	Scale 1 – 5; 1 = Poor; 5 = Excellent
X ₆	Planning skill of the project team leaders	Scale 1 – 5; 1 = Poor; 5 = Excellent
X ₇	Coordinating skill of the project team leaders	Scale 1 – 5; 1 = Poor; 5 = Excellent
X ₈	Project team leaders' adaptability to changes in the project plan	Scale 1 – 5; 1 = Poor; 5 = Excellent
OUTPUT (Performance Measure)		
Z ₁	Ahead time	Project status value > 0
Z ₂	On time	Project status value = 0
Z ₃	Behind time	Project status value < 0

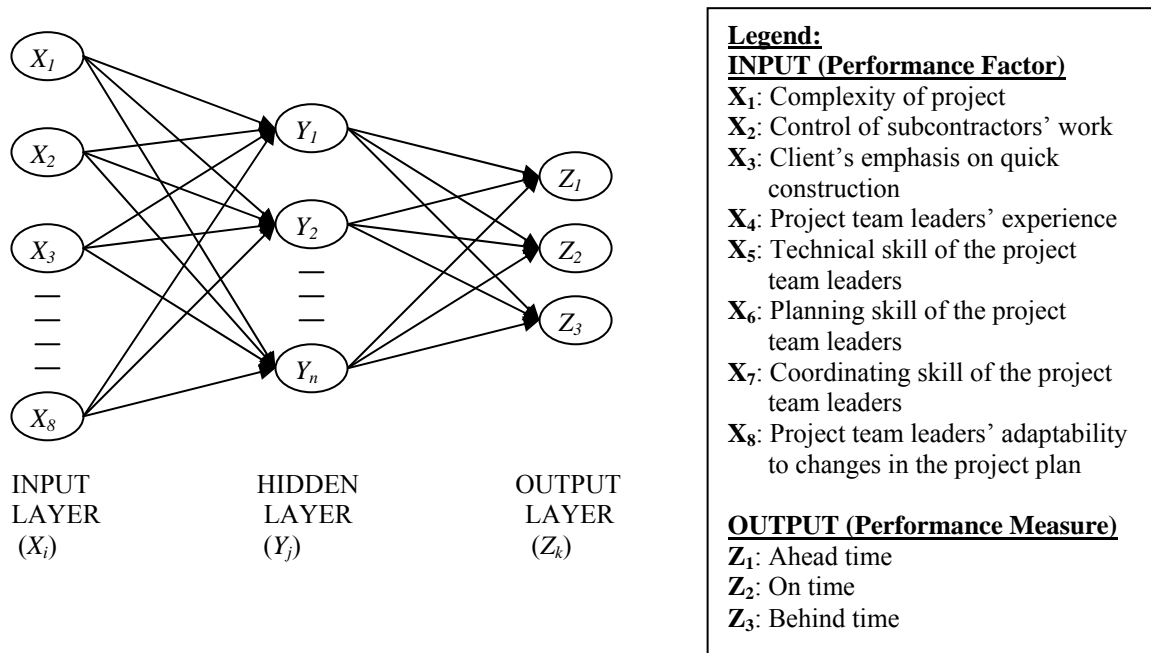


Figure 3: MLP Architecture Prediction Model

REFERENCES

- [1] Abd. Majid, M. Z. (1997), "Non-excusable Delays In Constructions." Ph.D. Thesis. Loughbrough University of Technology, UK.
- [2] Albert P. C. Chan, David Scott, Ada P. L. Chan. (2004). "Factors Affecting the Success of a Construction Project." Journal of Construction Engineering and Management. 2004;130(1):153.
- [3] Boussabaine, A. H. and Elhag, T. M. S. (1999). "Tender Price Estimation Using ANN Methods." Research Report No. 3. Construction Cost Engineering Group, School of Architecture & Building Engineering, University of Liverpool.
- [4] Cheung, S. O., Tam, C. N. And Harris, F. C. (2000). "Project Dispute Resolution Satisfaction Classification through Neural Network." Journal of Management in Engineering. ASCE. 16(1): 70-79.
- [5] Mark Konchar, Victor Sanvido. (1998). "Comparison of U.S. Project Delivery Systems." Journal of Construction Engineering and Management. 1998;124(6):435-44.
- [6] Ogunlana, O. S., Bhokda, S. and Pinnemitr, N. (2001). "Application of Artificial Neural Network (ANN) to Forecast Construction Cost of Buildings at the Pre-design Stage." Journal of Financial Management of Property and Construction. 6(3): 179-192.

The Effect of Cement and Water Cement Ratio on Concrete Paving Block

Tung-Chai Ling¹, Hasanan Md Nor^{1*}, Rachmat Mudiyono¹

¹ Department of Geotechnics and Transportation
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5531580, Fax: +60-7-5566157, E-mail: tcling611@yahoo.com

Abstract

This paper presents experimental results regarding the effects of cement content and water cement ratio (w/c) in the production of concrete paving block. Laboratory trials were conducted to compare and investigate the effect of 12 % and 15 % cement content in concrete mixtures at five different w/c ratios. In total, 60 concrete mixtures were cast and tested to determine dry density, compressive strength and skid resistance of concrete paving block samples. The design strength level ranging from 20 to 33 MPa and 31 to 42 MPa was achieved using w/c ratio from 0.50 to 0.70 for 12 % cement content and 0.45 to 0.65 for 15 % cement content, respectively. This study shows that, it was feasible to produce paving blocks prepared with 12 % cement that satisfied the minimum compressive strength (30MPa) in accordance to MA20 for trafficked area less than 3 tonnes gross weight.

Keywords: Concrete Paving Block; Cement, Water Cement Ratio, Skid Resistance

1. Introduction

In various countries, the concrete block pavement (CBP) becomes an attractive engineering and economical alternative to the both flexible and rigid pavement [1]. The strength, durability and aesthetically pleasing surface of paver [2] have made CBP ideal for many commercial, municipal and industrial applications, such as parking area, pedestrian, traffic intersection, container yards, etc.

In conventional, concrete paving block are produce based on a mixture of Portland cement, natural fine and coarse aggregate. The cement in the concrete mix is the finest material, and therefore has the greatest surface area per unit mass, and thus the greatest water requirement per unit mass. The high water requirement of the cement is also moderated by much lower water requirement of the aggregate, as typically most paving mixes have aggregate: cement ratio ranging between 5 and 7 [3].

For any given cement content there is optimal water content ratio. Using more water in the mix assist in reducing macroscopic entrapped voids, but too much water increases microscopic capillary voids. Conversely using less water has the potential to allow a closer packing of cement particles, but makes it so much more difficult to expel the air voids, as clearly less water means reduces lubrication/mobility.

Thus, the purpose of laboratory mix design in this study is to optimize the water content for 12 % and 15 % cement content in each mixture. Optimizing the water content

optimizes the overall strength of the concrete, and has potential to produce a hardened concrete paving block of the required minimum quality.

2. Experimental Work

2.1 Material Properties

The materials used to develop concrete mixture in this study consist of natural aggregate, additive and ordinary Portland cement complying with MS 522. The natural aggregates used include natural river sand as the fine aggregate having a maximum particle size of 4 mm and fineness modulus of 2.62 and crushed granite with nominal size less than 10 mm and 5.84 fineness modulus as the coarse aggregate. The weight ratio of coarse to fine aggregate of all paving blocks was kept to about 1: 2 throughout the whole experimental works.

2.2 Concrete Mixture

Two series of mixes were prepared using coarse and fine aggregate, cement, water and additive. The difference between the two series was the cement to aggregate and sand ratio. Where, (cement: aggregate: sand) 12 %: 29.33 %: 58.67 % and 15 %: 28.33 %: 56.67 % were used in series I and II, respectively.

A total of five w/c ratios designated ranging from 0.50 to 0.70 and 0.45 to 0.65 with cement contents of 242 kg/m³ and 298 kg/m³ were prepared in series I and II, respectively.

* Corresponding Author. E-mail: hasanan@fka.utm.my,
Tel: +60-7-5531500, Fax: +60-7-5566157

The mixed materials used were approximately 8.5 kg for each batch of three paving blocks samples.

2.3 Fabrication and Curing of Test Pavers

The paving blocks were fabricated in steel moulds with internal dimensions of 200 mm in length, 100 mm in width and 60 mm in depth as shown in Figure 1. The mix was poured into the mould in two layers of about equal depth. Compaction was applied manually using a hammer at each layer. The concrete blocks were then removed from the steel moulds one day after casting and cured in air at room temperature (Figure 2) for 7 and 28 days until tested.

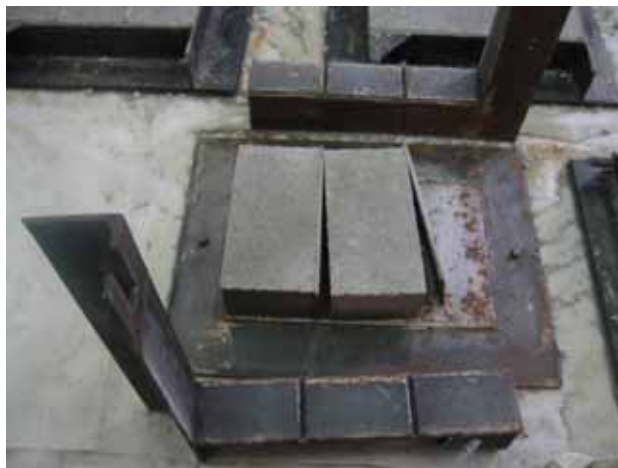


Figure 1: Fabrication and demoulding of concrete paving block specimens



Figure 2: Air curing at room temperature

2.4 Test Methods

A range of tests were carried out at the Structure and Material Laboratory, in Universiti Teknologi Malaysia to determine dry density, skid resistance and compressive strength at 7 and 28 days of the paving blocks specimens. The skid resistance of paving block was determined using a British Pendulum Skid Resistance Tester (Figure 3) and it was expressed as the measured British Pendulum Number (BPN) as specified by ASTM E303-93 [4].



Figure 3: Skid resistance test

The compressive strength was determined using a compressive testing as shown in Figure 4. The load, increased at a rate of 0.30 kN/s, was applied to the nominal area of block specimen. Prior to the loading test, the block specimens were soft capped with two pieces of plywood. The compressive strength was calculated by dividing the failure load by the loading area of the block specimen.



Figure 4: Compressive strength test

3. Experimental Results and Discussions

3.1 Dry Density

Results in Figure 5 indicated that dry density decreased to as low as about 2.09 g/cm³ at 0.50 w/c for 12 % cement content. For 15 % cement content the dry density of the concrete blocks ranged from 2.16 to 2.20 g/cm³ depending on the w/c ratio in the mixture. Because of high specific gravity of cement, unit weight of mixtures containing 15 % cement is slightly higher compare with 12 % cement content at the same w/c ratio. Moreover, decrease of w/c ratio in concrete mixture, which in turn reduces the unit weight of the mixtures.

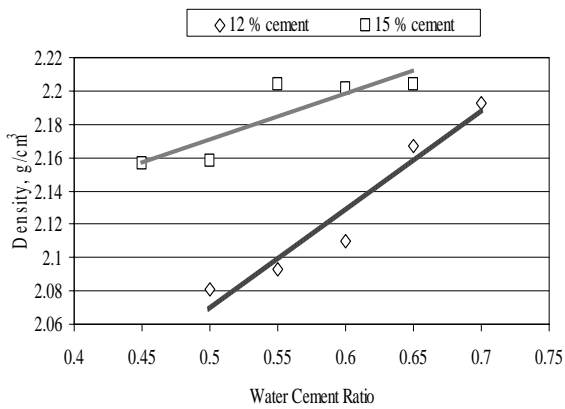


Figure 5: Dry density of paving blocks

3.2 Compressive Strength

The results of series I and II are summarized in Figure 6 and 7, respectively. Each presented value is an average of three samples measurement. The results presented in Figure 6 shows a systematic reduction in compressive strength with the decrease in w/c ratio for the paving blocks. The reason for the strength reduction could be attributed by insufficient water in the spaces between the cement grains to fully convert each individual grain of cement into gel. Therefore the inner cores of the cement particles remain unhydrated. But in series I specimens, it can be seen that the paving block specimens prepared at 0.65 and 0.70 w/c ratio satisfy the requirement of MA20 [5] with the average compressive strength higher than 30 MPa.

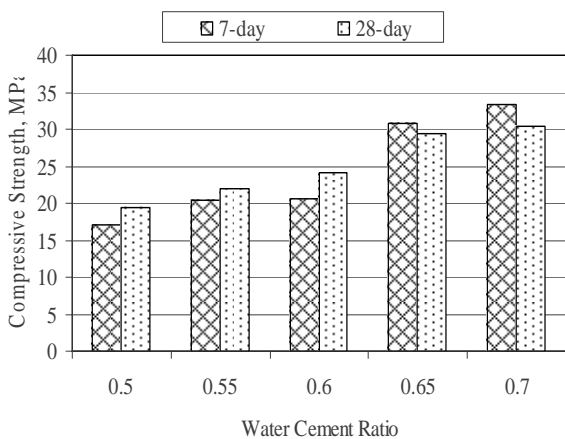


Figure 6: Compressive strength of series I paving blocks

The results presented in Figure 7 indicate that, the 7-day and 28-day compressive strength of series II paving blocks at five different w/c ratios. In the five w/c ratio, 0.55 w/c ratio gained the highest strength (42 MPa) compare to other w/c ratio. It is believed that at this w/c ratio, cement grains were close to each other to begin with the greatly increase in density of the microstructure and consequently compressive strength.

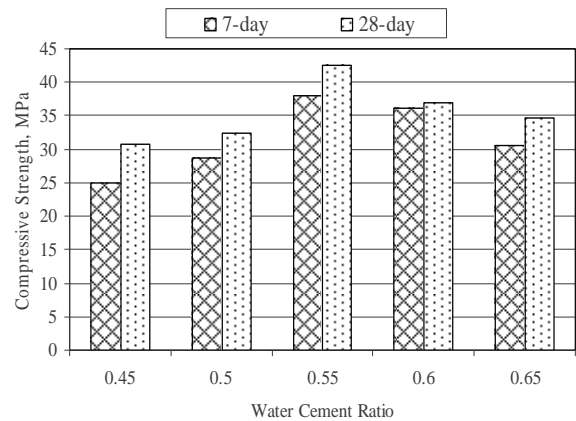


Figure 7: Compressive strength of series II paving blocks

3.3 Skid Resistance

Skid resistance was measured in accordance to ASTM E3030-93, four swings of the pendulum were made for each test surface paving block. The relationship between w/c ratio and BPN is shown in Figure 8. In general, the blocks produced in this study satisfy ASTM requirement that BPN were higher than 45. It is found that skid resistance was slightly higher for low w/c ratios. It was mainly due to the rough surface texture of paving blocks to create more friction as the pendulum passed across it.

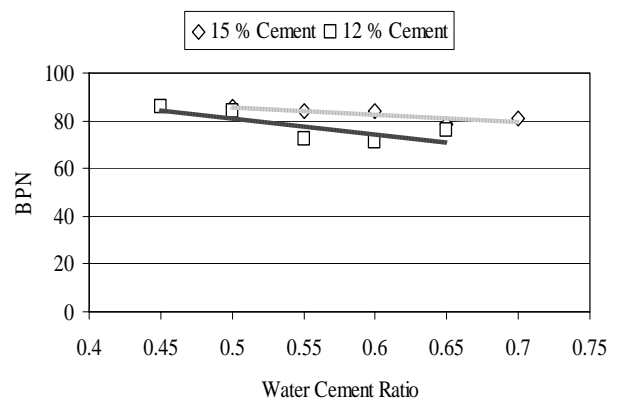


Figure 8: Skid resistance of paving block specimens

4. Conclusions

1. Dry density and compressive strength of concrete paving block is affected differently depending on cement content and w/c ratio. Clearly, if the cement content increases relative to the optimum of water in the concrete mixture, resulting in better dry density and compressive strength.
2. The test results indicated that there was an optimum w/c ratio (0.70 and 0.55) for every cement contents (12% and 15%), respectively.
3. It is possible to fabricate paving block containing 12% cement at 0.70 w/c ratio that satisfied the minimum compressive strength (30MPa) in

accordance to MA20 for trafficked area less than 3 tonnes gross weight.

4. Concrete paving block at low w/c ratios is found to provide better skid resistance.

References

- [1] Proceedings, First International Conference on Concrete Block Paving, University of Newcastle-upon Tyne, U. K., September 1980.
- [2] Gonzalo, R.R., Smith, D. R., Miller, J. S. and Witczalc, M. W. 1990. Structural Design of Concrete Block Pavements. *J. Transp. Eng.*, ASCE 116 (5): 615-635.
- [3] Papenfus, N. 2003. Applying Concrete Technology to Abrasion Resistance. *In Proceeding of the 7th International Conference on Concrete Block Paving*, Sun City, South Africa.
- [4] ASTM E 303. 2003. Measuring Surface Frictional Properties Using British Pendulum Tester. American Standard for Testing and Materials.
- [5] MA 20. 1996. Specification for Concrete Segmental Paving Units. Concrete Masonry Association of Australia.

Development of Sections Properties Tables for Cruciform Column Using Universal Beam Sections

M.M.Tahir, P.N. Shek, C.S. Tan, S. Anis

Steel Technology Centre, Faculty of Civil Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5531616, Fax: +60-7-5576841, E-mail: mahmoodmdtahir@yahoo.com

Abstract

Steel columns are usually designed using I-sections which are designated as universal columns or universal beams sections. However, due to weak axis of the I-sections, the design of the columns is braced on the minor axis. In order to solve the bracing problem, a new prototype column has been developed using universal beam section. In this paper, the new prototype section called cruciform column has been introduced for multi-storey unbraced frames. The section properties for the cruciform column was developed based on the design guide by the Steel Construction Institute in accordance with BS 5950-1: 2000[1]. The dimensions and properties listed in the table are important values that are used in the design of steel structure. From the study, it can be concluded that it is possible to develop the section properties table based on the design guide established by the Steel Construction Institute and the proposed calculated method.

Keywords: cruciform column, compound member, compression resistance, effective length, universal beam

1. Introduction

The usage of compound member is not new in engineering construction. In Malaysia, the cruciform column was recently used in the design of a 28 storey high of office building in Jalan Tuanku Abdul Rahman for Maju Perdana Holding. However, the use of compound member such as cruciform column is not popular although some of its advantages are quite significant. These advantages need to be highlighted so that the design engineers can consider the use of cruciform column as one of the alternative method to design the column. Moreover, with the development of the section properties tables, the design process can be further enhanced. Therefore, the development and advantages of the section properties tables will be highlighted in this paper. Many researches have been carried out on compound members/built-up columns but imperfections have occurred on those sections [2]. Fabrication and connection have caused problems to the compound member since they have resulted in the increased of cost and time. However, these problems are no longer the main obstacles in using use cruciform columns as most of the cutting and welding are precisely done with the use of an automated state of the art laser machine in the fabrication shop. As a result, cost and time of fabrication can be reduced.

Cruciform column is a combination of two universal beams where one beam is cut at mid-length and attached to the other beam by means of a fillet weld as shown in Fig. 1. This fillet weld is stronger than the parent materials that are

welded together. In order to achieve this strength the size of an effective weld is 0.7 multiplied by the size of weld. This should be greater than the thickness of the welded column web. The weld is usually welded on both sides in order to form a symmetrical section. As a result, a cruciform shape is formed and it has the same moment of inertia in the x-axis and y-axis which is very important to increase both the axial capacity and the stiffness of the column. The selection of using universal beam to form a cruciform column is intentional suggested by the authors due to some important advantages. Some of the advantages of using universal beam as cruciform column are listed as follows:-

1.1 Advantages of using universal beam as cruciform column.

1.1.1 Increase the axial capacity.

The axial capacity of column is usually calculated based on the area and compression resistance. The compression resistance depends on the effective length of the column and the radius of gyration. The division of effective length of the column with the radius of gyration is called λ (λ). This λ value will determine the axial load capacity of the cruciform column according to the code of practice[3]. The greater the value of λ the lesser will be the compression resistance of the column. The shape of universal beam section which gives higher moment of inertia in major axis is the reason why it has been selected to be used in the formation of cruciform column instead of universal column section. As a result, higher axial load capacity can be achieved due to a higher value of λ . Moreover, the weight of

UB is lighter than the weight of UC section with the same moment of inertia.

1.1.2 Increase the stiffness of the column

The stiffness of column is defined as the moment of inertia of column divided by the effective length of the column section (I/L). The higher the moment of inertia the greater will be the stiffness. By inspection on the geometrical aspects of the universal beam sections, the major axis of moment of inertia is greater than the universal column sections for the same mass. As a result, the advantage of using universal beam is quite significant. The importance of having higher stiffness value is crucial in the design of unbraced frame as the increase in the stiffening of the frame will reduce the sway of the frame. The limit of sway of a multi-storey building is one of the important aspects of the design that needs to be satisfied. Column is one of the main contributions in the stiffening of the frame besides rigid connections and stiffness of beams in unbraced frame[4].

1.1.3 Saving in steel weight.

Universal beams of the same mass as universal columns have higher moment of inertia on both axes. By using universal beams as cruciform columns, the slenderness ratio can be reduced due to higher radius of gyration. Slenderness ratio is defined as the length of column divided by the radius of gyration (L/r)[5]. The radius of gyration is a function of moment of inertia and area (A) given in the formula as $r = \left(\frac{I}{A} \right)^{1/2}$. Higher value of moment of inertia will give rise to a higher radius of gyration[5]. According to the code of practice[1], by reducing the value of slenderness, λ , the compressive resistance of a member will increase.

1.1.4 Easy to fabricate and erect.

The shape of the universal beam with a greater depth compared to the universal column will enhance the cutting and welding process to form the cruciform section. The open space that develops between the flanges is wide enough for the fixing of bolts during erection as shown in Fig. 1. This is not the case if a universal column section is used for cruciform column as there would be almost no spacing between the flanges left for the installation of bolts during erection. Moreover, due to the limited depth of the column, the flange of the column overlaps to each other as shown in Fig. 2 for most of universal column sections.

Steel Construction Institute[1] has established the design concept and procedure for developing the properties of hot-rolled steel sections including member sizes, second moment of inertia, plastic modulus, and elastic modulus. The same concept was adopted to calculate the section properties for cruciform column. However, some of the calculated values which are not mentioned by the design guide [1] were proposed by the authors. The sections for cruciform column presented in this paper are only limited to hot-rolled universal beam section. The section properties tables are presented in such a way that the dimensions and properties listed are quite similar to the typical universal beam and

column section properties tables as used in the design. The tables start with the largest beam size to the smallest beam available for British sections which ranged from UB 1016×305×487 to UB127×76×26. The notation used to identify the cruciform columns is designated as CCUB (Cruciform Column Universal Beam). Tables 1(a and b) to tables 2(a and b) show the calculated values for the proposed CCUB sections. The development of dimensions and properties tables is explained in accordance to the typical values used in the design of steel structure.

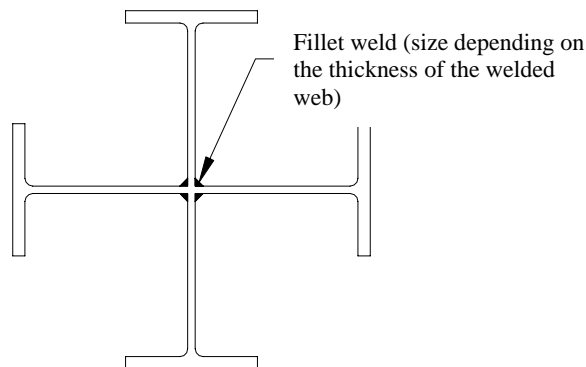


Fig. 1 Typical Drawing for Cruciform Column Using Universal Beam Section (CCUB 686x254x340)

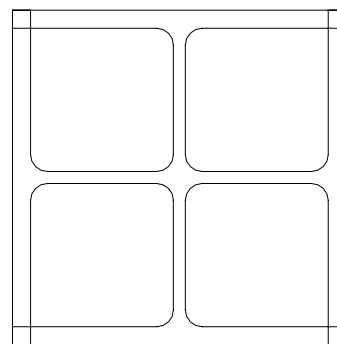


Fig. 2 Typical Drawing for Cruciform Column Using Universal Column Section (CCUC)

2. Development of dimensions and properties tables

Tables 1(a) to 2(a) are tabulated starting with the section designation followed by mass per meter, depth of section, width of section, thickness of web and flange, root radius, depth between fillets and ratio of local buckling of flange and web. Tables 1(b) to 2(b) are the continuation of the tables where the values of area of section, radius of gyration for x-axis and y-axis, second moment of area of x-axis and y-axis, elastic modulus of x-axis and y-axis and plastic modulus of x-axis and y-axis are presented. All these tabulated values are developed based on the design guide established by the Steel Construction Institute as explained earlier. However, some of the values were calculated by the proposed formula given by the authors to suit the geometrical configurations of CCUB sections as will be explained later in the paper.

2.1 Section designation (CCUB)

The section designation was based on serial size given by the section properties of British sections. The designation of first value is the depth, followed by the width and the weight of the CCUB section. For example, 1016×305×974 is designated as having a depth of 1016 mm, width of 305mm and weight of 974kg/m. The depth is the same as described for a typical UB section. The width is the width of the adopted UB section. The actual width is the same as the depth of the section. The weight was given as twice the weight of the typical UB section adopted.

2.2 Mass per meter

The value of mass in kilogram per meter was calculated based on the volume of the section per meter multiplied by the mass of steel taken as 7850 kg/m³. The volume of the section was calculated as the cross sectional area multiplied by one meter run. Therefore the mass of 973.4kg/m for 1016×305×974CCUB section was calculated as $1240 \times 10^{-4} \times 7850$ which is equal to 973.4kg per meter.

2.3 Depth, width, thickness of web and flange and root radius.

These dimensions are exactly the same values as tabulated in a typical UB section. The width is the same as the width of the flange of UB section. This is intentionally done in order to represent the size of an actual UB section that is used to be converted to CCUB section. Although the UB section is cut first before welding, the amount of loss during the process of cutting is ignored as the contribution to the calculation is not that significant. Therefore, the depth of the CCUB section is maintained as the same depth of the adopted UB section.

2.4 Depth between fillets

The depth between fillets was taken as half of the actual depth between fillets of the UB sections minus half of the web thickness at the center and the root radius of the fillet weld. The reduction of depth between fillets into half will improve the classification of the web from the slender section to a semi-compact or compact section. As a result, the compression resistance of cruciform column will increase as no reduction of the compression resistance [6].

2.5 Ratio for local buckling for flange and web

The ratio of local buckling for flange b/T is the same as the actual UB sections. However, due to the welded section which is attached to the centre of the beam, the ratio of local buckling for the web d/t was taken as a new depth between fillets. As a result, the axial capacity of the CCUB section will also improve.

2.6 Area of section

The area of section was taken as twice the area of typical UB section. The area of welded section to form the cruciform section is ignored as the area of steel lost during the process of cutting of the beam is assumed to be replaced by the welded section. The assumption of taking the area of the CCUB as twice the area of typical UB section was intentionally done so as to simplify the calculation process.

2.7 Radius of gyration of x-axis and y-axis

The radius of gyration of both axes x-x axis and y-y axis was expected to be the same. It was calculated based on the formula given as the root radius of second moment of area divided by the cross sectional area $\sqrt{\frac{I}{A}}$, which gives the

same value of x-axis and y-axis. The radius of gyration for adopted UB section has a smaller value for y-axis as compared to the CCUB section. The advantage of having a greater value of radius of gyration is quite significant especially in the determination of compression resistance.

2.8 Second Moment of Area (I)

The second moment of area of the section, often referred to as a moment of inertia, was calculated based on the calculated values given in the UB section properties tables. The basic formula to calculate the second moment of area is $I = BH^3/12 + AR$ where B is the width of the section, H is the depth of the section and R is the distance from the centre of the cross-sectional area of the designated area to the centre of the axis. The second moment of area of CCUB is calculated by the summation of the second moment of area of the x-x axis with the second moment of area of the y-y axis of the designated UB section. The calculation did not take into account the effect of welding as it is not that significant in the contribution to the calculation of the second moment of area.

2.9 Elastic Modulus (Z)

The elastic modulus is used to calculate the elastic moment capacity based on the design strength of the section or the stress at the extreme fiber of the section from a known moment. It is derived from the formula given as $Z = I / y$ where y is the distance to the extreme fiber of the section from the elastic neutral axis and I is the second moment of inertia of the CCUB section

2.10 Plastic Modulus (S)

The plastic modulus for cruciform column is derived from the basic formula used in the UB section as $S = BH^2/4$. The formula to calculate plastic modulus for cruciform column is then proposed by the authors as $S = B [D^2 - (D-2T)^2]/4 + t (D-2T)^2/4 + (D-2T) B^2/4 + (D-2T) t^2/4$ where:

B is the width of the section

D is the depth of the section

T is the thickness of the flange

t is the thickness of the web

3. Application of CCUB section to the steel construction.

As mentioned earlier, the use of CCUB section will increase the axial load capacity and the stiffness capacity of the section. Research on the design of multi-storey unbraced frame has shown the height of the building was limited up to 4 storey due to the problem of the sway of the building during maximum wind in conjunction with minimum gravity

load[7][8][9]. This is due to the sway of the unbraced frame on minor axis where the moment of inertia is very small compared to the major axis. By introducing the CCUB section, the moment of inertia can be increased which will then increase the stiffness of column of the multi-storey unbraced frame. The increment is not only on the minor axis but also on the major axis as well. As result, the limited height of the multi-storied unbraced frame can be increased to more than 4 storeys. However, more work needs to be done before any final conclusion can be made on the height of the building using CCUB section. The use of CCUB also solves the problem of connection on the minor axis. The connection of minor axis is usually done by end-plate connected to the web of the column or using thick plate welded across the flanges and connection to the beam. These types of connections are difficult to quantify in terms of the stiffness and the moment capacity of the connection[7]. Several researches have shown that the stiffness of the connections is less stiff compare to the major axis connection[7][8]. By introducing the CCUB section, the connection can be more predictable as the connection is the same as a typical major axis connection except that the column web is stiffen by connected sections at the centre. The connection is more stiff compared to the connection on minor axis[7], thus, the stiffening of the unbraced multi-storey frame can be further improved.

4. Discussion of results

Table 1 (a and b) to Table 2 (a and b) show the section properties for cruciform column using universal beam section. From the tables, the similarity in both major axis and minor axis values is very clear. The increment in section properties include the second moment of inertia (I), radius of gyration (r), gross cross sectional area (A), plastic modulus (S) and elastic modulus (Z). These increments will directly affect the behavior of the sections hence their performance. Some of the geometrical configurations of the cruciform column have the same values for both major axis and minor axis. These findings will indirectly indicate that the axial load capacity of cruciform column has the same capacity regardless of the orientation of the axis. On the other hand, the increment in section capacity means a reduction in steel weight. However, this not the case for UB sections as the minor axis always govern the design of the axial load capacity. This can be explained by comparing the capacity provided by different hot-rolled sections with cruciform column as described earlier. From the comparison, cruciform column require less steel weight when dealing with the same loading. Therefore cost reduction can be achieved by using cruciform column as less steel is need. The actual cost saving however, can really be understood by considering the cost of fabrication to form the cruciform shape and the consideration of less steel needed in the design of cruciform section. This scope of study however, is not discussed in this paper.

5. Conclusions

Steelwork designer should take into consideration the

advantages of cruciform columns in the design. Advantages such as increase in axial load capacity increase in stiffness, saving in steel weight, and easy to fabricate, are the main factors that need to be highlighted to the designers and fabricators. The section dimensions and properties shown in this paper have covered all the required values needed in the design of steel members. It can be concluded that the adoption of Steel Construction Institute method and the proposed method introduced by the authors has successfully established the dimensions and properties tables of cruciform column CCUB sections using UB sections.

References

- [1] The Steel Construction Institute and The British Constructional Steelwork Association Limited (2000) Steelwork Design Guide to BS 5950: Part 1: 2000 Volume 1 Section Properties Member Capacities (6th Edition) Incorporating Amendment 1 (June 2002).
- [2] D. A. Nethercot and R M Lawson (1992) Lateral Stability of Steel Beams and Columns – Common Cases of Restraint, The Steel Construction Institute.
- [3] British Standards Institution (2000) British Standard Institute BS 5950: Structural Use of Steelwork in Building Part 1: Code of practice for design Rolled and welded sections, British Standards Institution, London.
- [4] N. S. Trahair, M.A. Bradford and D. A. Nethercot (2001) The Behaviour and Design of Steel Structures to BS 5950, Third Edition-British, Spon Press.
- [5] David A. Nethercot (1991) Limit States Design of Structural Steelwork, Second edition, Chapman & Hall.
- [6] You Chul Kim, Hidetoshi Tahara, Eiji Nakaji and Kohsuke Horikawa (1999) Buckling Characteristics of High Manganese Non-magnetic Steel and Carbon Steel Hybrid Cruciform Columns, Trans. JWRI, Vol. 28.
- [7] M. Md. Tahir (1997) "Structural and Economic Aspects of the use of semi-rigid Joints in Steel Frames", Ph.D Thesis, University of Warwick, United Kingdom.
- [8] Kim, Y.W. (1988) "The behaviour of beam-to-column web connections with flush end plates", M.Phil Thesis, University of Warwick.
- [9] P. R. Salter, G. H. Couchman, D. Anderson (1999) Wind-moment Design of Low Rise Frames, The Steel Construction Institute.

Acknowledgments

The authors would like to acknowledge the contribution of IRPA (vote 74019 for providing the fund to carry out the analysis and data gathering). The authors would also like to acknowledge the Steel Technology Centre of UTM for its cooperation.

Table 1a. Tables of dimensions and properties of CCUB sections for size 1016x305x794 to 457x191x134

Section Designation	Mass Per Metre	Depth of Section	Width of Section	Thickness		Root Radius	Depth Between Fillets	Ratio for Local Buckling	
				Web	Flange			Flange	Web
	kg/m	D mm	B mm	t mm	T mm	r mm	d mm	b/T	d/t
1016x305x974	973.2	1036.1	308.5	30.0	54.1	30.0	389.0	2.85	12.97
1016x305x874	873.8	1025.9	305.4	26.9	49.0	30.0	390.5	3.12	14.52
1016x305x786	785.4	1016.0	303.0	24.4	43.9	30.0	391.9	3.45	16.06
1016x305x698	698.8	1008.1	302.0	21.1	40.0	30.0	393.5	3.77	18.65
1016x305x628	628.6	1000.0	300.0	19.1	35.9	30.0	394.6	4.18	20.66
1016x305x544	544.6	990.1	300.0	16.5	31.0	30.0	395.8	4.84	23.99
1016x305x498	497.4	980.2	300.0	16.5	26.0	30.0	395.9	5.77	23.99
1016x305x444	444.0	970.3	300.0	16.0	21.1	30.0	396.1	7.11	24.75
914x419x776	776.0	921.0	420.5	21.4	36.6	24.1	365.0	5.74	17.06
914x419x686	686.6	911.8	418.5	19.4	32.0	24.1	366.0	6.54	18.87
914x305x578	578.2	926.6	307.7	19.5	32.0	19.1	383.4	4.81	19.66
914x305x506	506.8	918.4	305.5	17.3	27.9	19.1	384.5	5.47	22.22
914x305x448	448.4	910.4	304.1	15.9	23.9	19.1	385.2	6.36	24.22
914x305x402	401.8	903.0	303.3	15.1	20.2	19.1	385.6	7.51	25.53
838x292x453	453.0	850.9	293.8	16.1	26.8	17.8	355.0	5.48	22.05
838x292x388	387.6	840.7	292.4	14.7	21.7	17.8	355.7	6.74	24.20
838x292x352	351.8	834.9	291.7	14.0	18.8	17.8	356.1	7.76	25.43
762x267x394	393.6	769.8	268.0	15.6	25.4	16.5	318.7	5.28	20.43
762x267x346	346.0	762.2	266.7	14.3	21.6	16.5	319.4	6.17	22.33
762x267x294	293.8	754.0	265.2	12.8	17.5	16.5	320.1	7.58	25.01
762x267x268	267.8	750.0	264.4	12.0	15.5	16.5	320.5	8.53	26.71
686x254x340	340.4	692.9	255.8	14.5	23.7	15.2	285.1	5.40	19.66
686x254x304	304.8	687.5	254.5	13.2	21.0	15.2	285.8	6.06	21.65
686x254x280	280.2	683.5	253.7	12.4	19.0	15.2	286.2	6.68	23.08
686x254x250	250.4	677.9	253.0	11.7	16.2	15.2	286.5	7.81	24.49
610x305x476	476.2	635.8	311.4	18.4	31.4	16.5	244.3	4.96	13.28
610x305x358	358.0	620.2	307.1	14.1	23.6	16.5	246.5	6.51	17.48
610x305x298	298.4	612.4	304.8	11.8	19.7	16.5	247.6	7.74	20.98
610x229x280	279.8	617.2	230.2	13.1	22.1	12.7	254.6	5.21	19.43
610x229x250	250.2	612.2	229.0	11.9	19.6	12.7	255.2	5.84	21.44
610x229x226	226.0	607.6	228.2	11.1	17.3	12.7	255.6	6.60	23.02
610x229x202	202.4	602.6	227.6	10.5	14.8	12.7	255.9	7.69	24.37
533x210x244	244.0	544.5	211.9	12.7	21.3	12.7	219.2	4.97	17.26
533x210x218	218.0	539.5	210.8	11.6	18.8	12.7	219.8	5.61	18.94
533x210x202	202.0	536.7	210.0	10.8	17.4	12.7	220.2	6.03	20.38
533x210x184	184.2	533.1	209.3	10.1	15.6	12.7	220.5	6.71	21.83
533x210x166	165.6	528.3	208.8	9.6	13.2	12.7	220.8	7.91	22.99
457x191x196	196.6	467.2	192.8	11.4	19.6	10.2	187.9	4.92	16.48
457x191x178	178.6	463.4	191.9	10.5	17.7	10.2	188.4	5.42	17.94
457x191x164	164.0	460.0	191.3	9.9	16.0	10.2	188.7	5.98	19.06
457x191x148	148.6	457.0	190.4	9.0	14.5	10.2	189.1	6.57	21.01
457x191x134	134.2	453.4	189.9	8.5	12.7	10.2	189.4	7.48	22.28

Table 1a. Continuation of Tables of dimensions and properties of CCUB sections for size 1016x305x794 to 457x191x134

Section Designation	Area Of Section	Radius Of Gyration		Second Moment of Area		Elastic Modulus, Z		Plastic Modulus, S	
	A	Axis X-X	Axis Y-Y	Axis X-X	Axis Y-Y	Axis X-X	Axis Y-Y	Axis X-X	Axis Y-Y
	cm ²	cm	cm	cm ⁴	cm ⁴	cm ³	cm ³	cm ³	cm ³
1016x305x974	1240.0	29.05	29.05	1046700	1046700	5051	5051	25630	25630
1016x305x874	1114.0	28.95	28.95	933500	933500	4550	4550	22862	22862
1016x305x786	1000.0	28.78	28.78	828500	828500	4077	4077	20339	20339
1016x305x698	890.0	28.86	28.86	741500	741500	3678	3678	18166	18166
1016x305x628	800.0	28.73	28.73	660200	660200	3301	3301	16197	16197
1016x305x544	694.0	28.61	28.61	568000	568000	2868	2868	13931	13931
1016x305x498	634.0	27.88	27.88	492800	492800	2514	2514	12230	12230
1016x305x444	566.0	27.16	27.16	417550	417550	2152	2152	10463	10463
914x419x776	988.0	27.83	27.83	765040	765040	4153	4153	20789	20789
914x419x686	874.0	27.58	27.58	664960	664960	3646	3646	18150	18150
914x305x578	736.0	26.58	26.58	519800	519800	2805	2805	14033	14033
914x305x506	646.0	26.38	26.38	449600	449600	2448	2448	12175	12175
914x305x448	572.0	26.03	26.03	387640	387640	2129	2129	10560	10560
914x305x402	512.0	25.57	25.57	334723	334723	1853	1853	9196	9196
838x292x453	578.0	24.64	24.64	351060	351060	2063	2063	10256	10256
838x292x388	494.0	24.16	24.16	288266	288266	1714	1714	8503	8503
838x292x352	448.0	23.80	23.80	253799	253799	1520	1520	7539	7539
762x267x394	502.0	22.23	22.23	248175	248175	1612	1612	8039	8039
762x267x346	440.0	21.96	21.96	212150	212150	1392	1392	6919	6919
762x267x294	374.0	21.57	21.57	173955	173955	1154	1154	5717	5717
762x267x268	342.0	21.32	21.32	155488	155488	1037	1037	5129	5129
686x254x340	434.0	20.19	20.19	176930	176930	1277	1277	6377	6377
686x254x304	388.0	20.06	20.06	156184	156184	1136	1136	5645	5645
686x254x280	356.0	19.94	19.94	141483	141483	1035	1035	5131	5131
686x254x250	318.0	19.62	19.62	122383	122383	903	903	4471	4471
610x305x476	606.0	19.28	19.28	225340	225340	1772	1772	8991	8991
610x305x358	456.0	18.99	18.99	164410	164410	1325	1325	6623	6623
610x305x298	380.0	18.86	18.86	135208	135208	1104	1104	5463	5463
610x229x280	356.0	18.07	18.07	116305	116305	942	942	4713	4713
610x229x250	318.0	17.96	17.96	102542	102542	837	837	4171	4171
610x229x226	288.0	17.75	17.75	90754	90754	747	747	3710	3710
610x229x202	258.0	17.46	17.46	78695	78695	653	653	3241	3241
533x210x244	310.0	16.01	16.01	79428	79428	729	729	3660	3660
533x210x218	278.0	15.84	15.84	69763	69763	647	647	3229	3229
533x210x202	258.0	15.78	15.78	64212	64212	598	598	2976	2976
533x210x184	234.0	15.69	15.69	57619	57619	540	540	2680	2680
533x210x166	210.0	15.36	15.36	49547	49547	469	469	2324	2324
457x191x196	250.0	13.87	13.87	48077	48077	515	515	2592	2592
457x191x178	228.0	13.75	13.75	43109	43109	465	465	2332	2332
457x191x164	208.0	13.68	13.68	38921	38921	423	423	2116	2116
457x191x148	189.2	13.60	13.60	34991	34991	383	383	1905	1905
457x191x134	171.0	13.43	13.43	30832	30832	340	340	1689	1689

Table 2a. Tables of dimensions and properties of CCUB sections for size 457x152x164 to 127x76x26

Section Designation	Mass Per Metre	Depth of Section	Width of Section	Thickness		Root Radius	Depth Between Fillets	Ratio for Local Buckling	
				Web	Flange			Flange	Web
	kg/m	D mm	B mm	t mm	T mm	r mm	d mm	b/T	d/t
457x152x164	164.2	465.8	155.3	10.5	18.9	10.2	188.4	4.11	17.94
457x152x148	148.4	462.0	154.4	9.6	17.0	10.2	188.8	4.54	19.67
457x152x134	134.4	458.0	153.8	9.0	15.0	10.2	189.1	5.13	21.01
457x152x120	119.6	454.6	152.9	8.1	13.3	10.2	189.6	5.75	23.40
457x152x104	104.6	449.8	152.4	7.6	10.9	10.2	189.8	6.99	24.97
406x178x148	148.4	412.8	179.5	9.5	16.0	10.2	165.3	5.61	17.39
406x178x134	134.2	409.4	178.8	8.8	14.3	10.2	165.6	6.25	18.82
406x178x120	120.2	406.4	177.9	7.9	12.8	10.2	166.1	6.95	21.02
406x178x108	108.2	402.6	177.7	7.7	10.9	10.2	166.2	8.15	21.58
406x140x92	92.0	403.2	142.2	6.8	11.2	10.2	166.6	6.35	24.50
406x140x78	78.0	398.0	141.8	6.4	8.6	10.2	166.8	8.24	26.06
356x171x134	134.2	363.4	173.2	9.1	15.7	10.2	141.1	5.52	15.50
356x171x114	114.0	358.0	172.2	8.1	13.0	10.2	141.6	6.62	17.48
356x171x102	102.0	355.0	171.5	7.4	11.5	10.2	141.9	7.46	19.18
356x171x90	90.0	351.4	171.1	7.0	9.7	10.2	142.1	8.82	20.30
356x127x78	78.2	353.4	126.0	6.6	10.7	10.2	142.3	5.89	21.56
356x127x66	66.2	349.0	125.4	6.0	8.5	10.2	142.6	7.38	23.77
305x165x96	96.2	310.4	166.9	7.9	13.7	8.9	119.8	6.09	15.16
305x165x84	83.8	306.6	165.7	6.7	11.8	8.9	120.4	7.02	17.96
305x165x74	74.0	303.4	165.0	6.0	10.2	8.9	120.7	8.09	20.12
305x127x96	96.2	311.0	125.3	9.0	14.0	8.9	119.2	4.48	13.24
305x127x84	83.8	307.2	124.3	8.0	12.1	8.9	119.7	5.14	14.96
305x127x74	74.0	304.4	123.4	7.1	10.7	8.9	120.2	5.77	16.92
305x102x66	65.6	312.7	102.4	6.6	10.8	7.6	127.1	4.74	19.25
305x102x56	56.4	308.7	101.8	6.0	8.8	7.6	127.4	5.78	21.23
305x102x50	49.6	305.1	101.6	5.8	7.0	7.6	127.5	7.26	21.97
254x146x86	86.0	259.6	147.3	7.2	12.7	7.6	98.3	5.80	13.65
254x146x74	74.0	256.0	146.4	6.3	10.9	7.6	98.8	6.72	15.67
254x146x62	62.2	251.4	146.1	6.0	8.6	7.6	98.9	8.49	16.48
254x102x57	56.6	260.4	102.2	6.3	10.0	7.6	101.9	5.11	16.17
254x102x50	50.4	257.2	101.9	6.0	8.4	7.6	102.0	6.07	17.00
254x102x44	44.0	254.0	101.6	5.7	6.8	7.6	102.2	7.47	17.92
203x133x60	60.0	206.8	133.9	6.4	9.6	7.6	75.4	6.97	11.78
203x133x50	50.2	203.2	133.2	5.7	7.8	7.6	75.8	8.54	13.29
203x102x46	46.2	203.2	101.8	5.4	9.3	7.6	74.4	5.47	13.78
178x102x38	38.0	177.8	101.2	4.8	7.9	7.6	63.4	6.41	13.21
152x89x32	32.0	152.4	88.7	4.5	7.7	7.6	51.1	5.76	11.34
127x76x26	26.0	127.0	76.0	4.0	7.6	7.6	38.7	5.00	9.68

Table 2b. Continuation of Tables of dimensions and properties of CCUB sections for size 457x152x164 to 127x76x26

Section Designation	Area Of Section	Radius Of Gyration		Second Moment of Area		Elastic Modulus, Z		Plastic Modulus, S	
	A	Axis X-X	Axis Y-Y	Axis X-X	Axis Y-Y	Axis X-X	Axis Y-Y	Axis X-X	Axis Y-Y
	cm ²	cm	cm	cm ⁴	cm ⁴	cm ³	cm ³	cm ³	cm ³
457x152x164	210.0	13.41	13.41	37775	37775	405	405	2032	2032
457x152x148	189.0	13.36	13.36	33717	33717	365	365	1820	1820
457x152x134	171.2	13.20	13.20	29843	29843	326	326	1620	1620
457x152x120	152.4	13.14	13.14	26295	26295	289	289	1431	1431
457x152x104	133.2	12.86	12.86	22015	22015	245	245	1210	1210
406x178x148	189.0	12.36	12.36	28855	28855	350	350	1750	1750
406x178x134	171.0	12.26	12.26	25695	25695	314	314	1565	1565
406x178x120	153.0	12.21	12.21	22803	22803	281	281	1391	1391
406x178x108	138.0	11.96	11.96	19741	19741	245	245	1216	1216
406x140x92	117.2	11.77	11.77	16228	16228	201	201	988	988
406x140x78	99.4	11.40	11.40	12920	12920	162	162	797	797
356x171x134	171.0	11.03	11.03	20822	20822	286	286	1439	1439
356x171x114	145.2	10.87	10.87	17148	17148	239	239	1194	1194
356x171x102	129.8	10.79	10.79	15108	15108	213	213	1055	1055
356x171x90	114.6	10.60	10.60	12881	12881	183	183	906	906
356x127x78	99.6	10.28	10.28	10528	10528	149	149	732	732
356x127x66	84.2	10.06	10.06	8529	8529	122	122	598	598
305x165x96	137.6	9.63	9.63	12763	12763	206	206	1032	1032
305x165x84	117.4	9.59	9.59	10795	10795	176	176	876	876
305x165x74	102.6	9.50	9.50	9267	9267	153	153	755	755
305x127x96	122.4	9.06	9.06	10036	10036	161	161	817	817
305x127x84	106.8	8.97	8.97	8585	8585	140	140	702	702
305x127x74	94.4	8.92	8.92	7507	7507	123	123	615	615
305x102x66	83.6	8.95	8.95	6695	6695	107	107	533	533
305x102x56	71.8	8.77	8.77	5521	5521	89	89	444	444
305x102x50	63.2	8.51	8.51	4578	4578	75	75	373	373
254x146x86	109.6	8.12	8.12	7221	7221	139	139	701	701
254x146x74	94.4	8.04	8.04	6108	6108	119	119	597	597
254x146x62	79.4	7.82	7.82	4861	4861	97	97	481	481
254x102x57	72.2	7.61	7.61	4184	4184	80	80	402	402
254x102x50	64.0	7.46	7.46	3564	3564	69	69	345	345
254x102x44	56.0	7.27	7.27	2960	2960	58	58	290	290
203x133x60	76.4	6.55	6.55	3281	3281	79	79	398	398
203x133x50	64.0	6.43	6.43	2648	2648	65	65	324	324
203x102x46	58.8	6.21	6.21	2269	2269	56	56	279	279
178x102x38	48.6	5.54	5.54	1493	1493	42	42	209	209
152x89x32	40.6	4.77	4.77	923.8	923.8	30	30	151	151
127x76x26	33.0	4.00	4.00	528.7	528.7	21	21	104	104

Modeling an Appointment Scheduling for Vehicle Inspection Center

Nuzulha Khilwani Ibrahim, Habibollah Haron, Nurulhuda Firdaus Mohd Azmi

Faculty of Computer Science and Information System
Universiti Teknologi Malaysia 81310 UTM Skudai, Johor

Abstract

This paper presents a preliminary analysis of appointment scheduling for a series of Vehicle Inspection Center (VIC). Effective scheduling systems have the goal of matching demand with capacity so that resources are better utilized and customer waiting times are minimized. In this paper, appointment scheduling problems at VIC are explored, investigated and presented. We offer solutions using constraint logic programming techniques in this new application area, with aims to improve the inspection process and to achieve greater efficiency and throughput at the VIC.

Keywords:

Vehicle scheduling, constraint logic programming, appointment system.

1. Introduction

Scheduling is the problem of assigning a set of tasks to a set of resources subject to a set of constraints. Examples of scheduling constraints include deadlines (e.g., job i must be completed by time t), resource capacities (e.g., there are only four drills), precedence constraints on the order of tasks (e.g., a piece must be sanded before it is painted), and priorities on tasks (e.g., finish job j as soon as possible while meeting the other deadlines). Examples of scheduling domains include classical job-shop scheduling, manufacturing scheduling, and transportation scheduling.

Scheduling is the process of assigning tasks to a set of resources. It is an important concept in many areas such as computing and production processes. In mathematical terms, a scheduling problem is often solved as an optimisation problem, with the objective of maximizing a measure of schedule quality. For example, an airline might wish to minimize the number of airport gates required for its aircraft in order to reduce its operating costs. Scheduling is important in modern production and chemical industries, where it can have a major impact on the productivity of a process. Common objectives in this type of scheduling are to minimize the make span (duration) of production or to maximize total profit for a given set of customer demands. Modern computerised scheduling tools greatly outperform the manual (heuristic) scheduling methods commonly employed in the industry.

Companies use backward and forward scheduling to plan their human and material resources. Backward scheduling is planning the tasks from the due date to determine the start date and forward scheduling is planning the tasks from the start date to determine the shipping date or

the due date. Scheduling problems have to satisfy specific conditions, especially constraints concerning requirements of resources in limited number. Besides the constraints, all of which must be satisfied and which are called hard constraints, there are also constraints which should be satisfied as far as possible; these are termed soft constraints.

In the service industry, customers expect short waiting times and management desires high throughput. These are primary issues addressed in this research, where improvements in scheduling areas play important roles in satisfying both the clients and management. The first section of this paper begins by definition of scheduling. The literature review from past papers of scheduling application presented in the next section. It is then followed by a section which thoroughly discusses the background review of the case study. Problems occurring in the case study are then modelled preliminarily and shown in the section. Also, the expected output from this research described here. Finally, the last section offers conclusions of research work.

2.0 Various Solution

Automated scheduling is a relevant and current research field. For automated scheduling there are different methods and approaches. Constraint Logic Programming (CLP) with constraints over finite integer domain, CLP (FD), has been established as a practical tool for solving discrete combinatorial problems. A scheduling problem can be suitably modeled in terms of a set of constraints. However, to generate a solution, a search is necessary.

2.1 Constraint Programming

Georg [2002a] implement Asynchronous Constraint Solver (ACS) as a separated object in his project with a distributed time scheduling tool. ACS can integrate Dynamic Constraint Processing (DCP) into concurrent Object-oriented Programming (OOP). Georg [2002b] wrote : “constraints are autonomous objects that can be posted and retracted in concurrently running solvers that communicate via common constraint variables”. In this project, he used the Asynchronous Constraint Solver (ACS) paradigm to solve finite domain constraints. Georg [2001] implement a distributed time scheduling tool where appointments can be organized automatically with constraint satisfaction. Georg also wrote that, to solve Distributed Constraint Satisfaction Problem, it has to gain information from constraints to other constraints. With all the constraints, large parts of search space can be cut earlier.

2.2 Constraint Logic Programming

Alessandro et al. [2005] described a constraint-based solution to the protein folding problem on *face-centred cubic lattices* – a biologically meaningful approximation of the general protein folding problem. In the paper, they have presented and tested a new search heuristic (*Bounded Block Fails*), which is well suited for the protein folding problem. Dal [2003] proposed a well-founded combination of the constraint solvers used in the constraint programming languages CLP(SET) and CLP(FD) semantically. They developed a combined constraint solver and they showed how static analysis can help in organizing the distribution of constraints to the two constraint solvers.

Maria et al. [2000] said that independence for CLP languages is more complex than for logic programming as search space preservation is necessary but no longer sufficient for ensuring correctness and efficiency. Their study also yields new insights into independence in logic programming languages. In particular, they showed that search space preservation is not only a sufficient but also a necessary condition for ensuring correctness and efficiency of parallel execution. Denis et al. [1999] proposed an approach which utilizes geometric transformations to reduce the number of quadratic equations defining geometric constructions involving circles and straight lines.

Guy [1999] retraces the general evolution of Prolog IV, the novel paradigm from its roots in inference systems and optimisation, to its applications in model building and problem solving. Hans [1999] used CLP to develop a timetabling system for a university in Berlin. Instead of the application proved successfully, they are still studying the influence of different modelling techniques on the solution search and exploring techniques for heuristic solution search. Hubert et al. [1999] fixed their goal in the research is to survey the notion of constraint system and to give examples of constraint systems operating on various domains, such as natural, rational or real numbers, finite domains, and term domains. Emden [1997] showed the existing extreme fix point characterization of chaotic iteration to be applicable to prove convergence of Waltz filtering. In the paper, they have

presented a framework that can help the further development of language like Prolog IV. Panagiotis et al. [1994] have presented an extension of the parallel Constraint Logic Programming language ElipSys, that directed towards the development of multi-agent systems which have to deal with large combinatorial problems that are distributed in nature. They have remarked that their proposed framework is suitable for developing multi-agent systems in ElipSys, where the agents themselves may exploit the parallelism and the constraint satisfaction techniques offered by the language.

2.2 Designing Appointment Rules

The first decision level in Appointment Scheduling (AS) design, is to find the best appointment rule. Recently, there is a lot of interest and variable-interval appointment systems, and future research may continue to investigate other variations with multiple and variable blocks.

Tugba et al. [2003] wrote, the appointment rule used to schedule patients can be described in terms of three variables:

- i. Block-size (n_i) is the number of patients scheduled to the i th block. Patients can be called individually, in groups of constant size, or in variable block sizes.
- ii. Begin-block (n_1), also called the initial block, is the number of patients given an identical appointment time at the start of a session.
- iii. Appointment interval (a_i) is the interval between two successive appointment times, also called “job allowance”. Appointment intervals can be constant or variable. A common practice is to set equal to some function of the mean (and sometimes the standard deviation) of consultation times.

It is also stated that any combination of these three variables (n_i , n_1 , a_i) is a possible appointment rule. The following appointment rules have been adapted from Tugba et al. [2003].

- i. Rule 1 : Single-block (Figure 1)
- ii. Rule 2 : Individual-block with Fixed Interval (Figure 2)
- iii. Rule 3 : Individual-block with Fixed Interval and Initial Block (Figure 3)
- iv. Rule 4 : Individual-block with Variable Interval (Figure 4)
- v. Rule 5 : Multiple-block with Fixed Interval (Figure 5)
- vi. Rule 6 : Multiple-block with Fixed Interval and Initial Block (Figure 6)
- vii. Rule 7 : Variable-block (Figure 7)

Example for Rule 1:

There is only one block vehicles will be scheduled for the whole session.

Meaning that, the number of block is depending on the number of session . If there are 2 sessions for vehicles inspection in a day, there will be only 2 block vehicles will be scheduled.

Let : first session start at 8am, and second session start at 2pm. Number of vehicles per day is 100, number of session is 2.

Therefore : the block size will be 100 divided by two, equal to 50. it means, there will be 50 vehicles scheduled at the morning (maybe all 50 vehicles will come at 8am), and another 50 vehicles will be schedule at the afternoon (and also, maybe all 50 vehicles will come at 2pm).

Maybe you can imagine the situation at the waiting area; some vehicles that come later than the earlier vehicles cannot enter the parking area and all of them will queue up at the roadside; while the owners will stress with the situation. What they can do? Just follow the regulation made by the VIC.

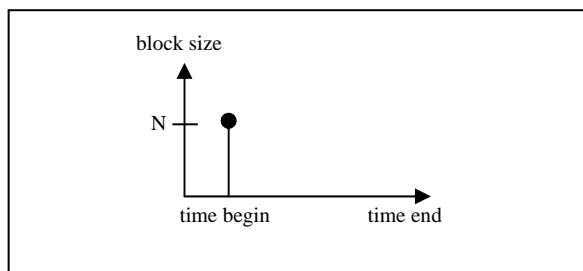


Figure 1 Allocation of Clients in Single-block Rule.

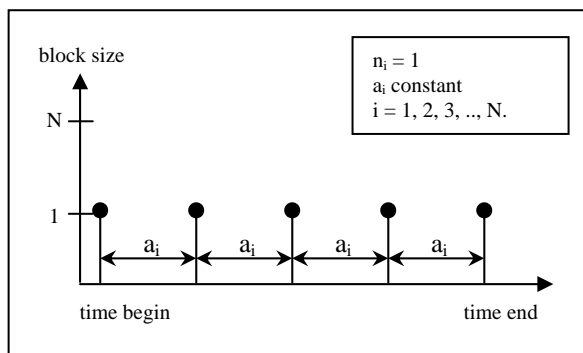


Figure 2 Allocation of Clients in Individual-block with Fixed Interval Rule.

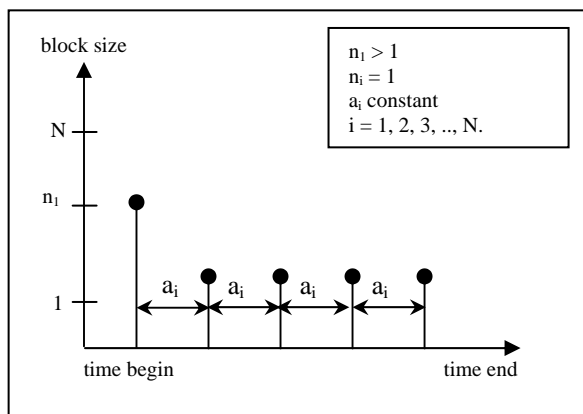


Figure 3 Allocation of Clients in Individual-block with Fixed Interval and Initial Block Rule.

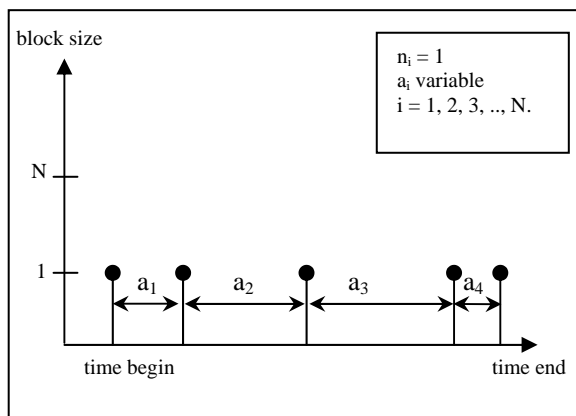


Figure 4 Allocation of Clients in Individual-block with Variable Interval Rule.

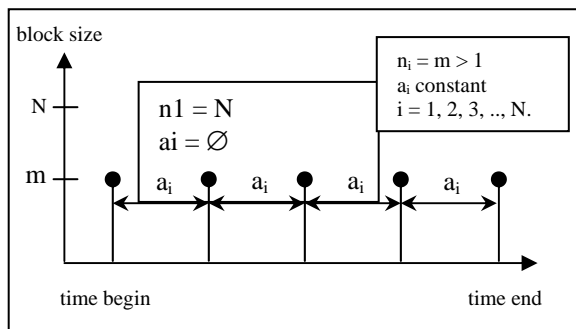


Figure 5 Allocation of Clients in Multiple-block with Fixed Interval Rule.

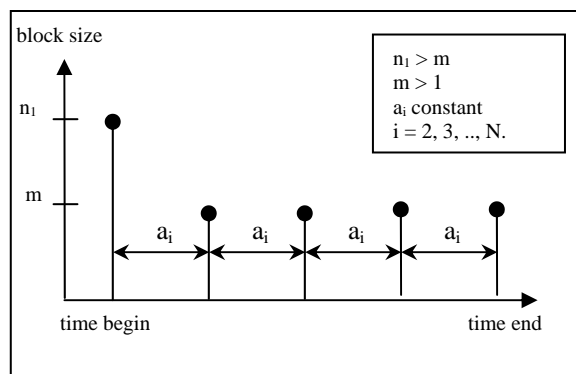


Figure 6 Allocation of Clients in Multiple-block with Fixed Interval and Initial Block Rule.

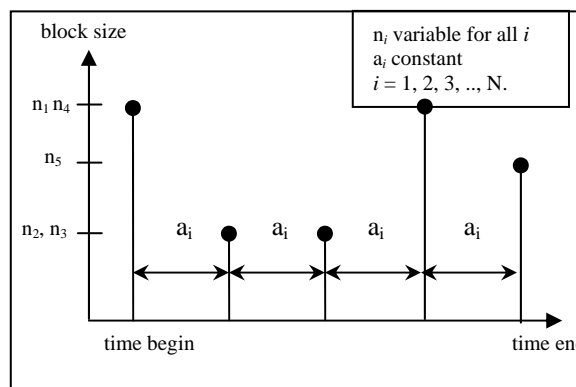


Figure 7 Allocation of Clients in Variable-block Rule.

3.0 Vehicle Inspection Centre

An exemplary application of a combined interactive and automatic inspection is scheduling system for the Vehicle Inspection Centre at twenty-eight centres nationwide, such as PUSPAKOM Wangsa Maju, Kuala Lumpur. This system will realize scheduling by means of CLP and a special algorithm for an efficient search. The output can be viewed as appointment notice as well as inspection schedule.

3.1 Problem Modelling

In the inspection of finance company's vehicle, there are three major inspections must be examine out of ten. There are some lanes for finance company' vehicle, depending on various centres nationwide. The important constraints are modelled as follows:

C_1 : The inspection activities can begin every 8 minutes and may run for different lengths of time as long as not exceeded 8 minutes time.

C_2 : There are restrictions with respect to the starting time for each inspection activity.

C_3 : Two staff about a same task must not be scheduled for the same lane.

C_4 : The staffs, vehicles, lanes, points, time slots, dates of each slot must not overlap.

C_5 : The number of inspections that a lane can conduct at a certain time is limited.

C_6 : Some inspection activities are only held during certain days where the following cases are possible : all days, A-Days (Full Day), B-Days (Half Day), C-Days (Friday), D-days (Holiday) either A-Days, B-Days, C-Days or D-Days.

C_7 : There are lanes about the same task carried out in parallel at the inspection area. Thus, a vehicle allocation can be chosen from among parallel lanes.

C_8 : Scheduling also involves allocating examiners to the individual inspection.

C_9 : Some inspections can only be held in certain specific lane, in some cases only one matches the requirements. These type of inspection is included adjustment cases for no-vehicles, emergencies, urgent client, second inspection and not on time.

C_{10} : The varying durations between the different points of the inspection must be taken into account.

C_{11} : The starting time preferences for inspection activities and the preferred lanes for examiners should also be taken into account if possible.

The constraint C_{11} is a soft constraint. The other constraints are hard constraints and must be satisfied.

3.2 The Output

There are two major output from this program. The two output are:

i. Appointment Notice

The first important output coming from this program is appointment notice or "Vehicle Inspection

Appointment". There are nine significant value must be suited to the notice. The notice must be issued for every vehicle before it can undergo for inspection. The sample of appointment notice is shown in **Figure 8**.

VEHICLE INSPECTION APPOINTMENT

Apmt. Id : _____

Vehicle Id : _____

Name of VIC : _____ Type of Inspection : _____

Apmt. Venue : _____ Apmt Lane : _____

Apmt. Date : _____ Apmt Time : _____

Figure 8 : Sample of Appointment Notice

ii. Inspection Schedule

The inspection schedule show the vehicle appointed for every slot in every lane. There are all lanes and slots information for everyday. Then, the schedule is verified by the assigned staff.

Figure 9a The Inspection Schedule (Booking)

Figure 9b The Confirmation for booked vehicle.

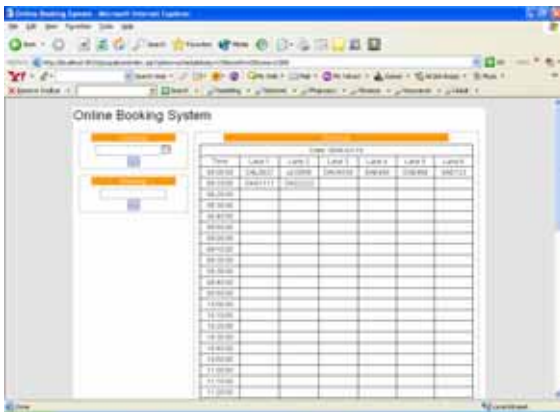


Figure 9c The Inspection Schedule (Confirmed)

iii. Database

The database will include the table of schedules as well as staff for the whole system. There are schedules and staff's information for everyday. Then, the table can be viewed by the admin to clarify the inspection data.



Figure 10a The Database View

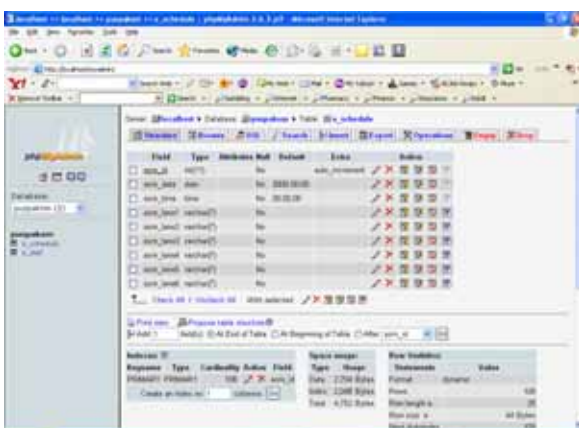


Figure 10b The Table (Schedule) Structure

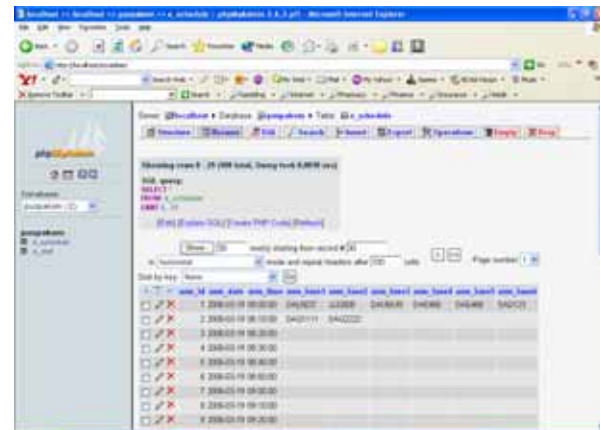


Figure 10c The Table (Schedule) Browse



Figure 10d The Table (Staff) Structure

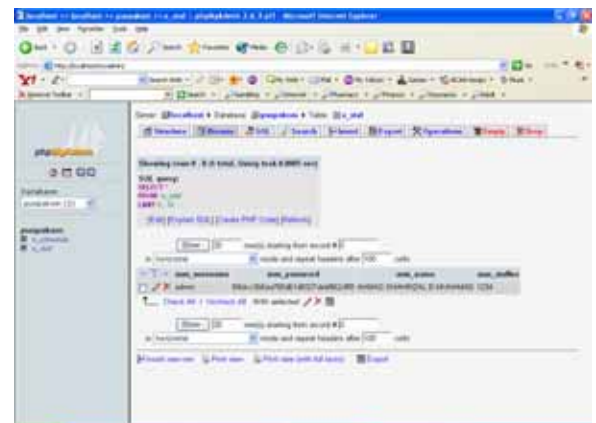


Figure 10e The Table (Staff) Browse

4.0 Discussion

This paper focuses on the interactive user intervention in the solution search via a graphical interface. The goals were to be realized by means of Constraint Logic Programming, as this approach is best suited to this study's needs. CLP combines the declarative of Logic Programming with the efficiency of constraint-solving algorithms.

There are a few contribution from this project, such as from application wise: it will aid the VIC to determine alternative strategies to improve its service. Then, the strategies can be implemented to its 28 centres. Appointment scheduling which is widely implemented in health care is now tailored to vehicle appointment scheduling.

REFERENCES

1. Alessandro, D. P., Agostino, D. and Enrico, P. [2005]. Heuristics, Optimizations, and Parallelism for Protein Structure Prediction in CLP(FD). International Conference on Principles and Practice of Declarative Programming. Proceedings of The 7th ACM SIGPLAN International Conference on Principles and Practice of Declarative Programming. 230-241.
2. Dal, P. [2003]. Integrating Finite Domain Constraints and CLP with Sets. International Conference on Principles and Practice of Declarative Programming. Proceedings of the 5th ACM SIGPLAN International Conference on Principles and Practice of Declarative Programming, Uppsala, 2003. 219-229
3. Denis, B., Laurent, T. and Jacques. C. [1999]. An Application of CLP: Checking the Correctness of Theorems in Geometry. Constraints: An International Journal, 4. Kluwer Academic Publishers, Boston, 1999. 383-405.
4. Emden, M. H. V. [1997]. Value Constraints in The CLP Scheme. Constraints: An International Journal 2, Kluwer Academic Publishers, 1997. 163-183.
5. Georg, R. [2001]. Distributed Constraint Satisfaction with Cooperating Asynchronous Solvers. Principles and Practice of Constraint Programming – CP 2001, 7th International Conference. CP 2001Paphos, Cyprus. Nov 26-Dec 1. Vol 2239/2001, 777.
6. Georg, R. [2002a]. Integrating Search Objects in Asynchronous Constraint Solving. Principles and Practice of Constraint Programming – CP 2002. 8th International Conference, CP 2002, Ithaca, New York, USA, Sept 9-13. Vol 2470/2002, 761.
7. Georg, R. [2002b]. Object-Oriented Constraint Programming with J.CP. MICAI 2002: Advances in Artificial Intelligence: Second Mexican International Conference on Artificial Intelligence Merida, Yucatan, Mexico, Apr 22-26. Vol 2313/2002, 194-203.
8. Guy, A. N. [1999]. From Prolog III to Prolog IV: The Logic of Constraint Programming Revisited. Constraints: An International Journal, 4. Kluwer academic Publishers, Boston, 1999. 313-335.
9. Hans-Joachim, G [1999]. University Timetabling Using Constraint Logic Programming. In G. Gupta (ed), Practical Aspects of Declarative Language, LNCS 1551, Springer-Verlag 1999. 320-334.
10. Hubert, C., Mehmet, D., Jean-Pierre, J. and Claude, K. [1999]. A Methodological View of Constraint Solving. Constraints: An International Journal, 4, Kluwer Academic Publishers, Boston. 337-361.
11. Maria, G. d. I. B., Manuel, H. and Kim, M. [2000]. Independence in CLP Languages. ACM Transactions on Programming Languages and Systems (TOPLAS), Volume 22, Issue 2 (March 2000). 296-339.
12. Panagiotis, S., Dimitris, M. and Constantin, H. [1994]. Extending A Parallel CLP language to Support the Development of Multi-Agent Systems. Symposium on Applied Computing, Proceedings of The 1994 ACM Symposium on Applied Computing, Phoenix, Arizona, United States. ACM Press New York, USA. 410-414.
13. Tugba, C. and Emre, V. [2003]. Outpatient Scheduling in Health Care: A Review of Literature. Production and Operations Management. Vol 12, No 4, 519-549.

Features Selection for Probabilistic Model of Response Generation in Dialogue Systems

Aida Mustapha, Md. Nasir Sulaiman, Ramlan Mahmod, Hasan Selamat

*Faculty of Computer Science and Information Technology
Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
Tel: +60-3-89466554, Fax: +60-3-89466577. E-mail: {aida, nasir, ramlan, hasan}@fsktm.upm.edu.my*

Abstract

The traditional, logic-based AI approach towards natural language generation (NLG) is basically knowledge- or rule-based and problem solving approaches are in the manner of top-down. The current research in NLG, however, are more focused on automatic data-driven techniques from machine learning and statistical NLG as empirical approach, whereby data is required to describe the language itself. Given the complexities of the dialogue structures in a mixed-initiative transaction dialogues, a probabilistic Bayesian learning model is proposed to model the knowledge acquired from the dialogue utterance analysis. The fundamental insight of the Bayesian model is to build multiple relationships for the user utterance input, in parallel to the server responses, compute the probability of each relationship, and choose the relationship between the server utterance and the corresponding user utterance that has the highest probability so the system can automatically learn its best response as to answer a particular user utterance.

In the probabilistic language modelling using Bayesian Network, the knowledge is introduced in two ways: within the structure of the network (relationships between variables) and the network parameters (the conditional probabilities distributions). In this research, the network structure for server utterances is being learned from SCHISMA dialogue corpus. Network interdependencies of the current server utterance are encoded on the previous user utterance based on dialogue acts, surface linguistic features and dialogue context features. The BN representation of the dialogue corpus is hoped to allow the system to reason about the dialogue utterances available so far and predict the next system response that will satisfy the user utterance both linguistically and contextually.

Keywords:

Dialogue System, Natural Language Generation, Corpus-Based Generation, Probabilistic Language Model, Bayesian Network.

1. Introduction

Many rule-based conversational agents or dialogue systems suffer from the same knowledge engineering bottleneck, which is the manual construction of rules to understand the user utterance and to generate response, a task of which are both time consuming and labour intensive. Also, the dialogue system operates only in a restricted and well-known domain. The ultimate way to train a dialogue system is by feeding the learning program with conversation transcripts (corpus) so the system can learn the pattern response rules automatically instead of specifying them manually. The research objective is how to train a dialogue system to jumpstart a conversation in a particular language and knowledge domain, given a sample or training corpus. This ability will allow users to transform the system behaviour only by change of corpora.

The study of probabilistic and decision-theoretic inference in the area of dialogue systems has been explored in a preliminary fashion for conversational games [1], dialogue act recognition [2], and for study in anaphora resolution [3]. Probabilistic approach is promising because the presence of uncertainties in dialogue utterances, and also knowledge representation that is very subtle and context-dependent. Based on this ground, a probabilistic model is in favoured

over a deterministic model in exploring automated response generation. The central problem of an automatic response generator has shifted from how to respond (generating surface structure of a response) to what to respond (generating the most accurate response) at a particular, specific turn in a conversation.

In this type of problem, there are many important pieces of information that are not directly available from the observation of surface utterance alone. However, with the current development of the standard Dialogue Acts Mark Up in Several Layers (DAMSL) annotation scheme [4], this knowledge are ready to be tapped on and available for learning. The availability of dialogue acts information can dramatically reduce the number of parameters to specify the Bayesian probabilistic model in response generation because the acts are the first level of semantic representation required in understanding the dialogue utterances.

2. The SCHISMA Corpus

The study of probabilistic response generator is using SCHISMA dialogue corpus [5], a theatre information and ticket reservation system. This system enable users to reserve a particular show from a wide range of available options, and both the user and the system must collaborate to achieve an agreement to several issues like the ticket price,

the seating arrangement or the discount availability. This type of negotiation process that takes place in the conversation is common in transaction dialogues, making the dialogue model far more complex than the usual question-answering systems. Also, at any point, both parties may request information from each other and the user particularly, may retract any previous decisions to proceed conversation in an opposite direction. Figure 1 illustrates an extract of dialogue in the corpus.

Table 1. Dialogue summary for SCHISMA

	User	System
# of dialogue	64	
# of utterance	920	1127
# of word	4582	15983

U What will be on in the theater next week (19 March)?
 S There is no show on that date.
 U And on 18 March?
 S In the period 18 March 1994 until 20 March 1994 you can go to Deelder Denkt and Indonesian Tales.
 U At what time does Deelder start?
 S The show starts at 20:00.
 U How much does it cost
 U and are there still places?
 S Do you have a reduction card?
 U No
 S The price for the show "Deelder Denkt" is f26,00.
 S And there are still 82 places free.

Figure 1. SCHISMA Dialogue Extract

SCHISMA also a mixed-initiative model, which consist two types of interactions: the inquiry and transaction [6]. During inquiry, the system is user-initiated because they will inquire about details of the shows like the dates, artists, reviews or authors, while the system will entertain to all the questions. When the conversation arrives at the point where the users indicate that they would like to make the reservations, the system will shift into transaction mode where the system now takes the initiative. Starting from this point onwards, the system will ask users series of questions like the number of tickets to reserve, the discount cards and others. User will answer the questions to complete the reservation details required by the system. This dialogue model is known as transaction dialogue. The content of dialogue corpus is summarized as in Table 1.

2.1. DAMSL Annotation Scheme

SCHISMA corpus is readily tagged using dialogue act annotation scheme based on DAMSL framework [5]. SCHISMA annotation consists of five layers, each of which covers different aspect of communicative functions. The five layers are information-level, forward-looking functions (FLF), backward-looking functions (BLF), topic management and surface features. This research concerns on two levels only, the forward-looking and backward-looking functions. Both levels indicate the communicative functions or the dialogue acts of an utterance. FLF tags indicate the type of speech act that the utterance is conveying, for example, *assert*, *info-request* and *commit*. BLF tags indicate

how the particular utterance relates to the previous utterance and include *answers* (*positive*, *negative* or *no-feedback*) to questions, degree of understanding or disagreement. Table 2 and 3 summarize the classes of forward-looking functions and backward-looking functions respectively.

Table 2: Dialogue acts (FLF) for SCHISMA

FLF	User	System
conventional	29	31
commit	0	4
offer	0	66
action_directive	239	11
open_option	0	111
query_if	71	38
query_ref	433	165
assert	123	694
exclamation	4	0
explicit_performative	2	0
other_ff	19	7
TOTAL	920	1127

Table 3: Dialogue acts (BLF) for SCHISMA

BLF	User	System
signal_understanding	7	0
signal_non_understanding	4	20
positive_answer	162	399
negative_answer	30	42
no_answer_feedback	3	63
correction_feedback	0	1
accept	70	54
reject	13	7
reject_part	2	1
hold	39	161
maybe	1	0
no_blf	589	379
TOTAL	920	1127

2.2. Dialogue Act Features

Dialogue act or intention recognition is not being carried out by the proposed automated response generator due to the availability of the dialogue act information from the DAMSL-annotated corpus. However, research in dialogue act recognitions (DAR) are also based on surface linguistic features of the dialogue utterance. We argue that in the absence of DA-annotated corpus, our response generator can be extended to include the DAR component, which can be run in tandem with analysis of our utterance type (UT) as described in the following section. Even though dialogue acts is important in classifying the specific intention of utterance, it is insufficient to perform interpretation of the utterance because the other component of equal importance is the content of the utterance.

There are important relationship between the surface linguistic features with the FLF and BLF as provided by the DA-annotated corpus. Further analysis to each utterance revealed that there are specific structural linguistic patterns from a particular combination of the forward-looking and backward-looking functions. The patterns are generalized

based on surface linguistic features only. These patterns are useful in determining the utterance type but they carry extra weight in context because are being preconditioned by the combination of FLF and BLF. The following section will discuss in detail the characteristics of surface linguistic features and the dialogue context features.

3. Surface Linguistic Features

Clustering and tagging utterances on the basis of surface features can be carried out using machine-learning techniques. The result will be classes of utterances described by a combination of surface properties and can be used to identify change of initiatives and determine the correct surface structure of the expected response. We use two kinds of surface features, which are the utterance type and content type for each utterance based on the information from the surface utterance.

3.1 Utterance Types

Utterance Type (UT) denotes the general structure of the server utterances, whether it is a declarative or imperative sentences or *wh*-question or yes/no questions. However, for utterances in SCHISMA corpus (being a human/machine dialogue), it also contains incomplete sentences like short answers, commands or remarks, consisting of single verb person (VP), noun person (NP) or prepositional phrases (PP). A separate class for text is used to categorize long utterance that contains multiple utterances, usually for server explaining the synopsis or the review of a performance. Feature classes are modified from the utterance types developed for dialogue act classifications in the SCHISMA corpus [7]. Similar to the previous experiment, cue patterns based on surface information are used to automatically classify the utterances. Examples of pattern cues are the position of verb, the presence of questions mark in *wh*-questions and others. The final classification for the utterance type for each user and server utterances is purely determined by the surface linguistic features in Table 4.

Table 4: Utterance types

Features	Description
dec	no <i>wh</i> -word, finite verb on 2 nd position
whq	<i>wh</i> -word 1 st position, finite verb on 2 nd position
ynq	finite verb on 1 st position, subject on 2 nd position
imp	finite verb on 1 st position, no subject
short	short answer i.e. prepositional phrases (PP), noun phrases (NP), proper names, adjectives, adverbs and numbers.
meta	Greetings i.e. thanks and bye, confirmation i.e. yes and negation i.e. no
text	long text i.e. reviews, list of genres, titles, authors, artists

The utterance type is very important cue to signal the server utterance of the structure of its expected response. For example, in choosing the best response for a question, the

server may exclude considering the probabilities of *wh*- and *yn*-questions, and look for either declarative type or imperative type of response utterance. This convention is basically determined by the grammatical structure of the sentence and the lexical type of the constituents. Another useful indication to classify the utterance type is by studying the combination of words and collocation [8]. For example, *please* or *would you* are a good cue for imperative utterance, and *are you* indicate that the utterance is a yes/no question.

3.2 Content Types

Content Type (CT) indicates the content intention of each utterance in the dialogue. Content intentions can also be seen as semantic tags that are assigned to every utterance based on the keywords detection to indicate the intended information to be conveyed from the utterance. To understand the content of a query utterance, we analyze the choice of *wh*-word i.e. *what* for title, *which genre* for genre or *what time* as time. Combination of the utterance dialogue act and the content will provide the system with semantic of the user input, and lead to the expected answer in the correct response structure. For example, the *when* question indicates the expected response from server should be a date.

We proposed 5 classes of content features, which are strictly domain dependant to SCHISMA Ticket Reservation System. The content classes and the word classes it contained are summarized in Table 5. The first column lists the content classes, and the second column lists the word classes. For example, consider the following user utterance, "*What will be on in the theatre next week (19 March)?*". Cue word *what* indicates that the user is asking for some information about the performance, and the given word class *date=19 March* indicates that the user is setting a parameter on a specific date. This combination of information will tag the user utterance with content type "*title*".

Table 5: Content types

Features	Description
performance	title, genre, artist, group, time, day, date, review
reservation	ticket, total_ticket, cost, total_cost, avail, reduction
theater	room, rank, balcony, floor, address
person	author, director, composer
other	utterance contains none of the above cue words

Content types are important to determine the semantic input for a given user utterance, based on the word classes mentioned in the utterance and other structural features of the utterance.

4. Dialogue Context Features

Dialogue analysis for context features is made possible by the presence of two most important dialogue acts according to DAMSL layers; the forward-looking functions (FLF) and backward-looking functions (BLF). Unlike the surface linguistic features that are determined by the surface structure

of the utterance alone, context features classify each utterance according to its positioning and time of presence during the conversation. The determination is based on theory of Dialogue Games and Negotiation Games with the advantage of DA-annotated corpus that provides the dialogue acts.

A dialogue game is a metaphor presented by [9] that is used in the study of exchange patterns in dialogues. During the game (exchanging utterance), each user and server plays a specific role and expects the opponent would do the same too. In case of mixed-initiative corpus like SCHIMA, both parties must respect the party that is currently holding the initiative, and must know when it is time for the initiative to change sides. The analysis of dialogue game will provide the semantic of moves in the conversation. We argue that if the system can learn its expected role from a given user utterance, it can find the highest probability response that satisfy the input utterance at the most correct timing.

The idea for negotiation game is adapted for dialogue systems based on ideas by [10]. It is an important base for SCHISMA analysis because in ticket reservation system, both participants must collaborate together to achieve an agreement before a reservation for a particular ticket show is reserved. Given the dialogue acts, the different phases during the negotiation process can be captured. [6] has presented a rough characterization of the dialogue acts that occur in each phase. The dialogue acts discussed are not specific to DAMSL data annotation scheme. However, it provides a strong basis to understand the contribution of different types of dialogue acts during the negotiation.

Table 6: Dialogue context features

Features	Description
initiative-type	inquiry, transaction, neutral
negotiation-type	open, inform, propose, confirm, close

Within our scope of research, we argue that the relationships between input and response utterance can be better postulated if we have the information of which phase the system is currently in, so it can limit its response choices and return the most accurate response. This feature is important because knowledge about the initiatives alone is insufficient because sequence of speaker turns does not necessarily consecutive. This implies that initiatives may change at any time and does not always alternate regularly. The basic principle in response generation is that, with respect to the content, the response utterance must be coherent and relevant to the input utterance. We propose two dialogue context features, the initiative-type and the negotiation-type. Table 6 describes both types of the dialogue context features.

4.1 Initiative Types

Any dialogue utterance in general can be an initiative, a response or neutral. An initiative requires a particular response from the opposing party, where the response can either be a direct answer or a signal of acceptance. For SCHISMA, initiative is an inquiry for information and the response consists of positive or negative reply; or else change of initiative into transaction. Before going into transaction, both parties must agree on a specific issue and commit to proceed to the next level. During transaction, the

exchange pattern is still the same as in inquiry state where queries are responded with assertives and assertions are followed by acknowledgements or confirmations. Only the lead is passed to system as the new control party of the conversation. We believe that it is therefore possible to formulate relationship between the dialogue acts and the exchange patterns within the context of the conversation, by identifying the party who is currently holding the right to initiate.

Initiative types denotes the mode of the system and serve as the indicator to be system if it should be in inquiry mode to answer questions or if it should be in transaction mode to ask questions. The system decisions may be learned from the corpus itself without specifying the dialogue model explicitly using the dialogue manager. Inquiry state can be defined as a series of information exchange between two participants where the inquirer has a certain information need. The goal of the user during this stage is to ask the system about the domain in order to satisfy his information need. During this mode, the role of the system to provide all information as per inquired by the users. However, in the process of feeding the information, the system may have questions to ask too, before concluding the answer. At the point where the system start asking, it is actually taking over the initiative to control the conversation and will shift into transaction state, while neutral initiative denotes the exchange of greetings.

4.2 Negotiation Types

To relate a particular server utterance with the input user utterance, the analysis of dialogue context should be extended to the identification of the general conversation phase at that particular utterance. A ticket reservation system like SCHISMA is a transaction dialogue because the user and the system must collaborate together to reach an agreement point of the details of particular reservation, before the reservation is confirmed and bound. The main activity during a transaction is negotiation, as described by [11] and [10] in the formal model of negotiation within the scope of dialogue systems. A negotiation can be characterized as an exchange of information that will eventually lead to agreement that satisfy both parties' information needs. The final agreement depends on the choices made during the negotiation process. Negotiation types are phases during this negotiation process. The phases are: opening, exchanging information, exchanging proposals, confirmation and closure as illustrated in Figure 2.

Usually in human-human conversation, people establish physical contacts and find a common ground of social level at the beginning of opening stage. In human-machine dialogues, both user and system exchange greetings to acknowledge presence of each other. During information phase, users usually seek all information and options by asking the relevant attributes, thus setting the negotiation space boundary. Within the context of SCHISMA corpus, user often asks about list of the performances like the titles, artists, authors and reviews. Typical dialogue acts that occur during this stage are questions (*queries*) and answers (*assertions*). Also, at this point users primarily hold the initiative and the system is in inquiry mode, which signals the system to answer questions only.

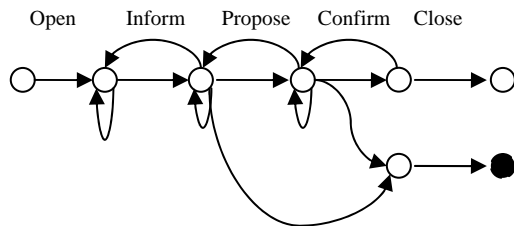


Figure 2. Phases in negotiation process

A proposal is identified when the negotiation space becomes restricted to certain topic. Proposals are different from information because it generates commitment if it is accepted. For example, if the user has agreed to a specific show, the following utterances are bound to the context only related to that particular show. Proposals may be offered from both parties. The details of the performances are like the date, time, cost of ticket, number of tickets and reductions if applicable. Dialogue acts in this phase are usually directives like *open_option* (server gives out options and user has to choose) or *action_directive* (user state their decision choice).

Following the phase is the confirmation phase, often marked with a question to confirm, and an answer is expected to either agree (*confirm*) or disagree (*disconfirm*). Until this point, all decisions are retractable. After this point, the user is bound to the transaction. In SCHISMA ticket reservation system, the confirmation concerns the total cost of the tickets. Therefore the system concludes the transaction by giving total cost and seeks confirmation from user to activate the reservation. After confirmation or disconfirmation, the transaction has to be concluded. Normally in this phase both parties exchange closing greetings and thanks.

5. Conclusion

The general problem for a probabilistic response generator is to choose the highest probability response that answers the user input accurately. Our task is therefore to compute the probability of each server utterance and return to user the response that has the highest probability. It is important that the machine learning techniques also learn other correlated parameters: is it an initiative or response? How does it relate its content to the context? What kind of dialogue acts it has? Is it assertive? Interrogative? Directives? Is the content positive or negative?

Constructing a probabilistic model for automatic response generation requires all dialogue acts, surface linguistic and context features as random variables in setting up the network. We argue that each server utterance is highly dependent on the variables that the joint probability distribution can give the system ability to learn its future response based on the flow of conversation so far. An automatic response generator does not have to rely on the pre-programmed pattern response rules to communicate with users but to rather learn its responses based on the training data. Table 7 summarizes the features being proposed for the experiment and the type of node it represent in constructing a Bayesian network.

Table 7. Surface and context features as random variables

Variables	Type	Description
UFLF	scalar	FLF of user input utterance
FLF	scalar	FLF of server utterance
BLF	scalar	BLF of server utterance
UtType	scalar	Utterance type of each utterance
IniType	scalar	Initiative type of each utterance
ConType	scalar	Content type of each utterance
NegoType	scalar	Negotiation type of each utterance

6. References

- [1] Pulman, S. 1996. *Conversational Games, Belief Revision and Bayesian Network*. In 7th Computational Linguistics in Netherlands (CLIN 96).
- [2] Keizer, S. 2001. *A Probabilistic Approach to Dialogue Act Clarification*. In Proceedings of Bi-Dialogue 2001.
- [3] Lemon, O., Parikh, P. and Peters, S. 2002. *Probabilistic Dialogue Modeling*. In Proceedings of 3rd SIGDial Workshop in Discourse and Dialogue pp125-128.
- [4] Allen, J. and Core, M. 1997. *Draft of DAMSL: Dialogue Act Markup in Several Layers*. The Multiparty Discourse Group, University of Rochester, Rochester, USA.
- [5] Hoeven, G.v.d., Andernach, T., Burgt, G.v.d., Kruijff, G., Nijholt, A., Schaake, A. and de Jong, F. 1994. *SCHISMA: A Natural Language Accessible Theater Information and Booking System*. In Proceedings of Speech and Language Engineering, Twente.
- [6] Hulstijn, J. 2000. *Dialogue Models for Inquiry and Transaction*. PhD Thesis. Univ. of Twente, Netherlands.
- [7] Andernach, T. 1996. *A Machine Learning Approach to the Classification of Dialogue Utterances*. In Proceedings of NEMLAP-2, Bilkent, Turkey.
- [8] Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Meteer, M. and Ess-Dykema, C. 2000. *Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech*. Computational Linguistics, 26:339-371.
- [9] Mann, W. 1988. *Dialogue Games: Conventions of Human Interaction*. In Argumentation, 2:511-532.
- [10] Di Eugenio, B., Jordan, P. W., Thomason, Richard, H., and Moore, J. D. 1998. *An Empirical Investigation of Proposals in Collaborative Dialogues*. In Proceedings of COLINGACL' 98, Montreal, Canada.
- [11] Walker, M. 1996. *Inferring Acceptance and Rejection in Dialog by Default Rules of Inference*. Language and Speech, 39(2-3):265-304.

Document Image Recognition Of Javanese Manuscripts

(Case Study of Javanese Manuscripts: Menak Sorangan I and Panji Sekar)

Anastasia Rita Widiarti ¹, Agus Harjoko ²

¹ Faculty of Mathematics and Natural Sciences
Sanata Dharma University, Paingan Maguwoharjo Depok Sleman, Yogyakarta 55002, Indonesia
Tel: +60-0274-883968, Fax: +60-0274-886529, E-mail:rita@staff.usd.ac.id

² Electronics And Instrumentation Lab.,
Faculty of Mathematics and Natural Sciences
Gadjah Mada University, Yogyakarta, Indonesia
E-mail: aharjoko@ugm.ac.id

Abstract

Yogyakarta hosts many collections of classic Javanese manuscripts that are invaluable cultural heritage. Converting these texts into digital format will give immense benefits, particularly in terms of text enhancement and preservation.

This research aimed to make a prototype of document image recognition system for Javanese manuscripts, which consisted of three steps. First, a given texts were read using optic devices, and saved as picture-type files *.jpg, which were then fed into the system as input data. The second step is to perform pixel-level processing on the input data to prepare them for further analysis. Such processing includes: binarization to reduce a gray scale to a binary image; orientation normalization to correct the skew of the document which typically results from improper paper feeding into the scanner; filling to reduce noise; thinning to get the skeleton of each character; and segmentation to separate document image into its components down to individual characters. The third step is character recognition to derive the meaning of the characters. Character recognition algorithms have two essential components: (i) feature extraction, i.e. by counting the number of pixels of the object in each unit of a character image, the characteristics of the image were determined. These characteristics were kept in the Javanese character database, which afterwards was used in the verification of a given character, and (ii) classification, i.e. by utilizing the modification of the Euclidean distance to match the Javanese character with its corresponding Latin character.

The research showed that the success rate of this system in recognizing the images of the given Javanese manuscripts, in this case Menak Sorangan I and Panji Sekar pages 3 and 4, was 81.28%. In conclusion, it can be said that the methods used in the stages of this document image recognition system has been appropriate.

Keywords: Character Recognition, Document Image Recognition, Pixel-level Processing

1. Introduction

Document images are visual representations of paper documents such as journals, facsimiles, business letters, and administrative forms. Document image recognition is an effort to transform document images into semantic representations [1]. Yogyakarta is famous for its rich cultural heritage, including art and literature. Among invaluable historical heritage Yogyakarta has is Javanese manuscripts, some of which are in the collections of Yogyakarta Kasultanan Palace and Pakualaman Palace [2]. The development of the document image recognition has offered a good opportunity for the preservation of these old texts found in Yogyakarta.

This article presents a software for document image recognition of manuscripts written in modern Javanese characters, outlining details of the pixel-level processing, the extraction stage to extract the distinctive characteristics of

each Javanese character, and the verification stage to verify the characteristics of the Javanese character.

2. Methods

O'Gorman and Kasturi [3] give a sequence of steps for document image analysis that can be modified for the image recognition of Javanese manuscripts. The first step is data capture, in which a given text is read using optic devices, and saved as picture-type files *.jpg. The second step is to perform pixel-level processing on the input data to prepare them for further analysis, and to create intermediary features to help character recognition. The third step is character recognition, by which several images of the same Javanese character, in different shapes and sizes, are translated into its corresponding Latin character.

2.1. Pixel Level Processing

The following step is the testing, which is initiated by pixel-level processing. This step employs the binarization process, followed by the orientation normalisation, filling, thinning, and finally the process of segmentation.

2.1.1 Binarization

Binarization is the first step in pixel level processing after data capture has been executed, which is a process to separate the object from its unwanted gray background [3]. This process is mainly intended to automatically determine the *threshold* value that will function to divide an image into two parts or groups, namely object group and background group.

The *threshold* of a given image $g(x,y)$, where $f(x,y)$ expresses the *gray-level* of a coordinate (x,y) , and T expresses a *threshold* value, is defined as follows [4]:

$$g(x,y) = \begin{cases} 1 & \text{jika } f(x,y) > T \\ 0 & \text{jika } f(x,y) \leq T \end{cases} \quad (1)$$

where 1 expresses the background and 0 expresses the object.

The approach used by Otsu method [5] is by executing discriminatory analysis, namely determining a variable that can differentiate the naturally emerging two or more groups. For example, to find the threshold value of an image with N number of pixels, and gray-level of 256. The probability for a pixel with gray-level i to appear is expressed as follows:

$$p_i = n_i / N \quad (2)$$

where n_i expresses the number of pixels with gray-level i . If k represents the threshold value which is to be determined from a black-and-white image, therefore, for a gray-level image, the value of k will range from 0 to L , where $L = 255$ expresses the highest gray level.

The threshold value of k can be calculated by maximizing the following equation:

$$\sigma_B^2(k^*) = \max_{1 \leq k < L} \sigma_B^2(k) \quad (3)$$

with:
$$\sigma_B^2(k) = \frac{[\mu_T \omega(k) - \mu(k)]^2}{\omega(k)[1 - \omega(k)]} \quad (4)$$

The value μ_T is the sum of multiplication of i times p_i . The value $\mu(k)$ is called the cumulative moment to the 0 toward the color intensity to the $-k$, and the value $\omega(k)$ expresses the cumulative moment value to-one toward the color intensity to the $-k$.

$$\omega(k) = \sum_{i=1}^k p_i \quad (5)$$

$$\mu(k) = \sum_{i=1}^k i p_i \quad (6)$$

$$\mu_T = \sum_{i=1}^L i p_i \quad (7)$$

2.1.2. Normalization of Orientation

Orientation normalization is executed to reduce the impact of imperfect orientation resulted from improper reading of image data using the scanner. It often happens that document images are not fed properly into the scanner, with the images not perpendicular to the main axis. Therefore, the image must be rotated to the skew angle θ , to correct the image orientation.

Rotation of an image with an angle being θ counter clock-wise (see Figure 1.) can be done using the following equation (8).

$$\begin{aligned} x' &= x \cos(\theta) - y \sin(\theta) \\ y' &= x \sin(\theta) + y \cos(\theta) \end{aligned} \quad (8)$$

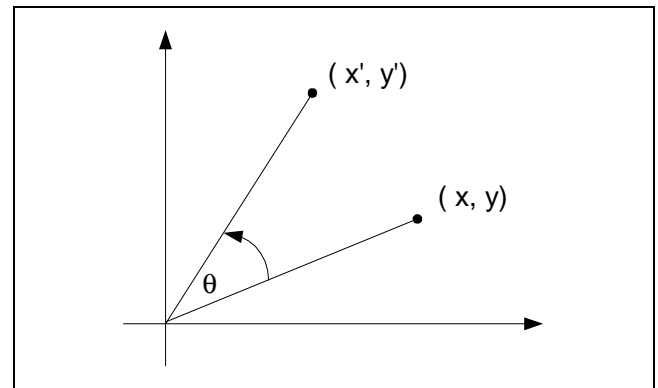


Figure 1. Rotation of an image with angle θ

To estimate the skew angle of the object to the main axis, can be done using moment orientation. The value θ can be determined using the following equation (9).

$$\theta = \frac{1}{2} \tan^{-1} \left(\frac{2\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}} \right) \quad (9)$$

where:

$$\mu_{p,q} = \sum_m \sum_n (m - \bar{m})^p (n - \bar{n})^q \quad (10)$$

expresses the central moment of ordo (p, q) , with \bar{m}, \bar{n} being the point of central moment which is defined as follows:

$$\bar{m} = \frac{1}{N} \sum_m \sum_n m, \quad \bar{n} = \frac{1}{N} \sum_m \sum_n n, \quad (11)$$

N expresses the number of pixels of an image, while m and n express the object coordinate of the image [6].

2.1.3. Filling

The next step in pixel processing is noise reduction step, which is also called *filling*. *Salt-and-pepper noise* is an example of *noise* that commonly appears as a result of bad

quality of document images, for example from improper copying of the document image.

Two important processes carried out in this step are dilatation and erosion, where the two processes are based on 8 neighbouring points from a given pixel. The erosion process aims to reduce the size of the foreground area, which has been resulted from the existence of foreground pixels in the background area, and the dilatation process is the counter process of erosion, that is adding foreground pixels to close the foreground area [3].

2.1.4. Thinning

Thinning or *medial-axis* is a process to reduce the image components to their essential information, or in other words, to get the skeleton of an image. Therefore, *thinning* can be called skeletonizing.

Hilditch algorithm [7] is one of the algorithms that can be used in skeletonizing. For example, eight neighboring pixels P2, P3, ..., P9 of a pixel P1 are illustrated below:

P9	P2	P3
P8	P1	P4
P7	P6	P5

Figure 2. Windows 3x3

To decide whether to delete pixel P1 or to keep it as part of the skeleton, the eight neighboring pixels are thus arranged as illustrated above, and two functions are needed:

1. B(P1), which expresses the number of neighboring pixels of pixel P1, which is the object point.
2. A(P1), which expresses the number of pattern 0,1 for the sequence of P2-P3-P4-P5-P6-P7-P8-P9-P2

Hilditch's algorithm consist of performing multiple passes on the pattern and on each pass, the algorithm checks all the pixel from black to white if it satisfies the following four conditions:

1. $2 \leq B(P1) \leq 6$
2. $A(P1) = 1$
3. $P2.P4.P8 = 0$ or $A(P2) \neq 1$
4. $P2.P4.P6 = 0$ or $A(P4) \neq 1$

This algoritma is no longer executed when there is no more point to change during one.

2.1.5. Segmentation

Segmentation is a process to detect individual regions or objects in the image. In document image analysis, segmentation occurs on two levels, first separating text and graphics, and second performing resegmentation of both text and graphics.

Segmentation process can be obtained from the projection profile [8]. For example, a binary image S with M

lines and N columns as illustrated in Figure 3. The vertical projection profile (P_v) of the image S is expressed as the number of back pixels perpendicular to y axis, which is equated below:

$$P_v[i] = \sum_{j=1}^M S[i, j]$$

(12)

And the horizontal projection profile (P_h) of the image S is the number of black pixels perpendicular to x axis:

$$P_h[i] = \sum_{j=1}^N S[i, j]$$

(13)

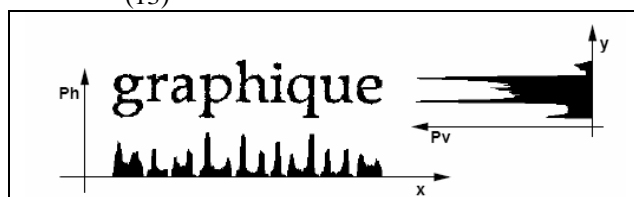


Figure 3. Profile of horizontal and vertical projection

2.2 Character Recognition

Character recognition aims to translate a sequence of characters that have various shapes and sizes, and it mainly works by developing algorithm for character identification. Character recognition can be complicated by the existence of various shapes, sizes, and skew angles of a single character.

Character recognition algorithm consists of two main parts, namely feature extraction and classification. The process of character recognition is aimed to extract the characteristics of a class of character by using algorithm of a given classification based on the extracted feature that is gained from the feature extraction process.

2.2.1. Feature Extraction

Generally the method carried out to extract the characteristics is as follows [9] :

1. A character image is divided into 3 x 3 parts so as to become 9 parts called unit.
2. From each unit in the complete image, the number of pixels of the object is counted.

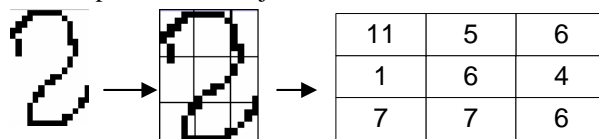


Figure 4. Illustration of the calculation of the characteristics of the Javanese character: tarung

2.2.2. Classification

The second step in the character recognition, and the last step in the program testing, is the classification process. In

this stage, an object is classified into a certain class based on its characteristics. In this research, classification is carried out with the natural approach as follows:

1. The value of each unit in a location corresponding to a character to be recognized is compared to the value of each character unit in the database. This comparison is carried out by making use of the modification of the Euclidean distance. When obtaining two character features a, b with 9 units, then the D_f distance (a, b) between these two features can be defined as follows:

$$D_f(a, b) = \sum_{i=1}^3 \sum_{j=1}^3 |a_{ij} - b_{ij}| \quad (14)$$

2. The least minimum distance of the results from comparison in step one is then obtained. The least minimum distance that is still under the permitted upper limit is then chosen as the character that best corresponds to the character under study.

3. Results

This research conducted the testing of the program. The output was afterwards compared to a manual reading/translation. The testing of the program was meant to know the reliability of this MATLAB-based application compared with results of manual reading/translation. The writer used the manual reading/translation by a librarian who graduated from the department of Javanese Literature and works specifically with Javanese literature.

From the document data of Menak Sorangan I page 3 as presented in Figure 2, with threshold being 20 [10] the outputs of the character recognition were obtained as seen in Table 2. The table shows that not all characters from the document image were correctly recognized. This can be seen in the Latin words that are marked with shading.

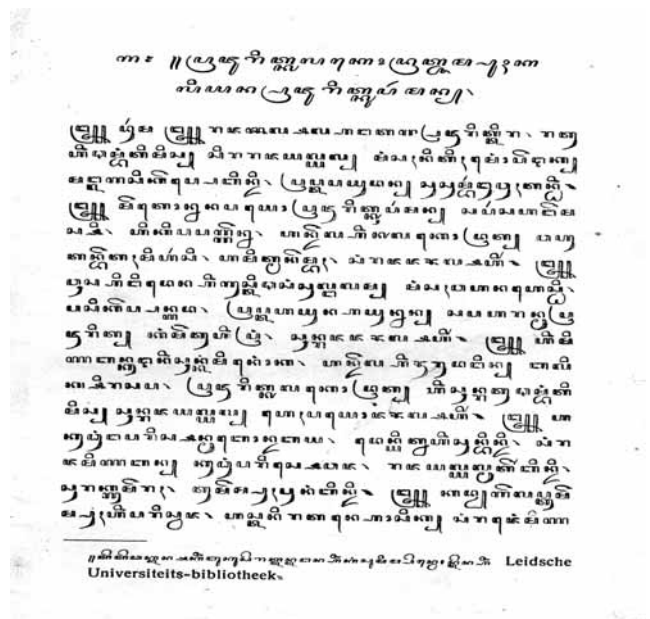


Figure 5. Menak Sorangan I document page 3

Table 1. Results of the Testing of the Program for Menak Sorangan I page 3

Line	Character recognition
1	1 : adeg2 adeg2 pra ju ri t ka la taling ka tarung drat ca ma puh ka
2	Li ya na pra ju rit ku par man PadaLingsa
3	PadaMadya Dur ma PadaMadya ra ja sa la sa la ha nga ta gal ga2 pra ju rit ti ra PadaLingsa ra tu
4	ing Tam bang ti mis si ra ra ja ya l ya l mang sah ni tih taling mang tarung pe thak
5	mang la ga si ke taling p pe bin di PadaLingsa prap ta pa yu dan su sum bar ngu suh tan ding PadaLungsi
6	PadaMadya me taling ta tarung na na pa taling ya pra ju rit ku par man sa par sa ha nge ma
7	sa si PadaLingsa hi ki sa sag gen na PadaLingsa han de la ing Ka la taling ka PadaLingsa drat pa yu
8	tan be tah mi yar si PadaLingsa ya mit ma nem bah PadaLingsa sang ra ja ja Ba la sa hir PadaLungsi PadaMadya
9	wu sa i ngi taling da na ing gus ti Ta sang sul nga la m mang sah pa ha na taling has thi PadaLingsa
10	pa si kep pan ga da PadaLingsa prap ta ha yu na ha yun nan sa sa : ha ran mu pra
11	jurit kang me tu ing pra PadaLingsa sun raja ja Ba lasa hir PadaLungsi PadaMadya ing mi
12	Na ban ku tha ning sun kang mi taling nang tarung ka PadaLingsa han de la ing Bu ru da ngin ba li
13	ka si ra sa sa PadaLingsa pra ju rit Ka la taling ka tarung drat ing sun ra tu Tam bang hi
14	mis sun ra ja ma l ya la1 la2 taling hah pa taling ya ja Ba la1 tarung sa hir PadaLungsi PadaMadya ha
15	ku dhung nga sa ri sa san mu taling ba tarungn ba ba ma tarung taling dan yit na ing sun bin di PadaLingsa sang ra
16	ja mi Na han ku dhung pa ri taling sa sa sa ja ,ra ja ya l dal yal mu ter bin di PadaLingsa
17	su rak gu me rah PadaLingsa tu me ma1 ma2 puh pu nang bin di PadaLungsi PadaMadya kad ya ge la1 ma2p tu me
18	ma puh ing pa ris wa ja PadaLungsi has ta ni ra ta taling na ha tarung sik sa ra taling jang mi Na
19	adeg2 adeg2 te te dhak a na sa king bu ku se rat tan ta nga na ing kang taling mi ma pe talingn wa n te na ing L ma2 i d tarung : h ma2
20	sa ma2, v na tarung t na i tarung h b i o t h na tarung k PadaLungsi

4. Discussion

Based on the character recognition as seen in Table 2., the table of the data output analysis of the character image recognition image is presented as follows:

Table 2. Output Analysis of the Document Image
Menak Sorangan I page 3

Line	Number of characters					Additional characters	
	Charac Ter	Recog Nized	Correct	Wrong	Diffe rent	Space	Other Character
1	18	18	18	0	0		
2	11	12	11	0	1	1	
3	22	23	22	0	1		1
4	25	25	25	0	0	1	
5	26	27	25	1	1		1
6	24	24	23	1	0	1	
7	23	27	20	3	4	2	2
8	25	25	24	1	0		
9	26	26	25	1	0		
10	23	24	21	2	1		1
11	21	21	20	1	0		
12	24	30	24	0	6	6	
13	23	27	21	2	4	1	
14	23	25	20	3	2		2
15	26	28	22	4	2	2	
16	24	25	21	3	1		1
17	24	26	23	1	2		2
18	26	27	25	1	1		1
19	26	26	20	6	0		
20	25	26	11	14	1	1	
Total	465	492	421	44	27	16	11

Table 3. shows that, from 465 characters in the document image, 421 characters have been successfully recognized. Thus the success rate of this document image recognition can be calculated as $(421/465) \times 100\%$, or 90.54%.

The same analysis on the output from the document image recognition of Menak Sorangan I page 3 earlier is afterwards applied on the analysis of the output from the other document image (see Table 2.). The summary of the analysis of program output is as follows:

Table 3. Summary of the Analysis of Program Output

Document title	Number of characters			Success rate
	Image	Recog nized	Correct	
Menak Sorangan I,	465	492	421	90.54

page 3				
Menak Sorangan I, page 4	526	562	434	82.51
Panji Sekar, page 3	477	509	345	72.33
Panji Sekar, page 4	538	577	429	79.74
Average of success rate				81.28

The result of the data analysis shows that the success rate of this document image recognition is 81.28%. The percentage shows that the methods used in this document image recognition of Menak Sorangan I and Panji Sekar have been relatively appropriate.

5. Conclusion

Based on the research, it can be concluded that the success rate of this system in recognizing the images of the given Javanese manuscripts, in this case Menak Sorangan I and Panji Sekar pages 3 and 4, was 81.28%, with threshold of characteristics distance being 20. The percentage of the success depends not only on the variety of character images stored in the database, but also on the quality of the paper, the quality of printing, and the complex and various forms of the printed Javanese character. The better the quality of the paper and the printing, the higher the success rate is. The more complex and various the form of a printed Javanese character, the more difficult it is to do document image recognition.

The results of this research show the necessity to study other methods of pixel-level processing and character recognition so as to get more optimal results. In addition, further researches should be conducted on other Javanese manuscripts, particularly those written in old Javanese characters.

Acknowledgments

I would like to express my gratitude to the Head of the Library of Sanata Dharma University, Dr. F. Susilo, S.J., for his permission to use the books in the collection of Pustaka Artati. I would also acknowledge the help of Yohanes Widodo, S.S. in translating the Javanese manuscripts to the latin characters.

References

- [1] Srihari, S.N.; Lam, S.W.; Govindaraju, V.; Srihari, R.K.; and Hull, J.J. 1986. *Document Image Understanding*. CEDAR: New York.
- [2] Suryakusuma, S. 2003. Kamus-Kamus Bahasa Jawa. http://www.tembi.org/perpus/2003_02_perpus03.htm

- [3] O'Gorman, L., and Kasturi, R. 1997. *Executive Briefing: Documen Image Analysis*. USA: IEEE Computer Society Press.
- [4] Gonzalez, R.C., and Woods, R.E. 1992. *Digital Image Processing*. USA: Addison-Wesley Publishing Company.
- [5] Otsu, N. 2003. A Threshold Selection Method from Gray-Level Histograms. www-users.itlabs.umn.edu/classes/Spring2003/csc8980/presentations/AthresholdSelectionMethodfromGray-Level.ppt.
- [6] Jain, A.K. 1989. *Fundamental of Digital Image Processing*. NJ: Prentice Hall Information and System Sciences Series.
- [7] Azar, D. 1997. Hilditch's Algorithm for Skeletonization. <http://jeff.cs.mcgill.ca/~godfried/teaching/projects97/azar/skeleton.html#algorithm>.
- [8] Zramdini, A., and Ingold, R. 1993. Optical font recognition from projection profiles. *Electronic Publishing* 6(3): 249-260.
- [9] Parker, J.R. 1997. *Algorithms For Image Processing And Computer Vision*. New York: Wiley Computer Publishing.
- [10] Widiarti, A.R. 2006. *Pengenalan Citra Dokumen Sastra Jawa: Konsep dan Implementasinya. Tesis S2 Ilmu Komputer*. Yogyakarta: Universitas Gadjah Mada.
- [11] Yasadipura I, R.Ng. 1936. *Menak Sorangan I*. Batawisentrem: Bale Poestaka.
- [12] Pakubuwono IV. 1933. *Panji Sekar*. Batawisentrem: Bale Poestaka.

Palmprint Verification Based on Fractal Codes and Fractal Dimensions

Darma Putra, I Ketut Gede¹, Adhi Susanto^{2*}, Agus Harjoko³, Thomas Sri Widodo^{2**}

¹ Departement of Electrical Engineering, Faculty of Engineering, Udayana University, Bali - Indonesia
email : ceit@telkom.net, phone +62-08155735367

² Departement of Electrical Engineering, Faculty of Engineering, Gadjah Mada University
Yogyakarta – Indonesia, phone* +62-08122780944, ** +62-08164265042

³ Electronic and Instrumentation Lab., FMIPA, Gadjah Mada University, Yogyakarta – Indonesia
email : aharjoko@ugm.ac.id, phone +62-08164267256

Abstract

This paper proposes a new technique to extract the palmprint features based on some fractal codes. The palmprint features representation is formed based on position of range blocks and direction between the position of range and domain blocks of fractal codes. Each palmprint representation is divided into a set n blocks and the mean value of each block are used to form the feature vector. The normalized correlation metrics are used to measure the degree of similarity of two feature vectors of palmprint images. We collected 1050 palmprint images, 5 samples from each of 210 persons. Experiment results show that our proposed method can achieve an acceptable accuracy rate with $FRR = 1.754$, and $FAR = 0.699$.

Keywords:

Digital image processing, Palmprint recognition, fractal dimensions.

1. Introduction

The personal verification becomes an important and highly demanded technique for security access systems in this information area. Traditional automatic personal recognition can be divided into two categories: token-based, such as a physical key, an ID card, and a passport, and knowledge-based, such as a password and a PIN. However these approaches have some limitations. In the token-based approach, the “token” can be easily stolen or lost. In the knowledge-based approach, the “knowledge” can be guessed or forgotten [21]. In order to reduce the security problem caused by traditional methods, biometric verification techniques have been intensively studied and developed to improve reliability of personal verification. Biometric-based approach use human physiological or behavioral features to identify a person. The most widely used biometric features are of the fingerprints and the most reliable are of the irises. However, it is very difficult to extract small minutiae features from unclear fingerprints and the iris input devices are very expensive [19]. Other biometric features such as of face, voice, hand geometries, and handwritten are less accurate. Faces and voices can be mimicked easily, hand geometries and handwritten can be faked easily.

Palmprints are the relatively new in physiological biometrics [18]. There are many unique features in a

palmprint image that can be used for personal recognition. Principal lines, wrinkles, ridges, minutiae points, singular points and texture are regarded as useful features for palmprint representations [21]. A palmprint has several advantages compared to other available features: low-resolution images can be used, low cost capture devices can be used, it is very difficult or impossible to fake palmprints, and their characteristics are stable and unique [18].

Recently, many verification/identification technologies using palmprint biometrics have been developed [2],[3],[4],[5],[11],[12],[13],[18],[21]. Zhang *et al.* [21] applied 2-D Gabor filter to obtain the texture features of palmprints. Pang *at al.* [13] used the pseudo-orthogonal moments to extract the features of palmprint. LI *et al.* [12] transformed the palmprint from spatial to frequency domain using Fourier transform and then computed ring and sector energy features. Connie *at al.*[2] extracted the texture feature of palmprint using PCA and ICA. Wu *et al.*[18] extracted line feature vectors (LFV) using the magnitudes and orientations of the gradient of the points on palm-lines. Kumar *et al.*[11] combined the palmprints and hand geometries for verification system. Each palmprint was divided into overlapping blocks and the standard deviation value of each block was used to form the feature vector.

In this paper, we propose a new technique to extract the features of palmprint based on fractal codes. This technique is different from the method we presented in [5].

2. Image Acquisition

All of palm images are captured using Sony DSC P72 digital camera with resolution of 640 x 480 pixels. Each persons was requested to put his/her left hand palm down on with a black background. There are some pegs on the board to control the hand orientation, translation, and stretching. A sample of the hand and pegs position on the black board is shown on Figure 1 (a).

3. Extraction of Palmprint

This research used the technique in [5] to extract the ROI of palmprint, consists of two steps in center of mass (centroid) method. These steps can be explained as follow. The gray level hand image is thresholded to obtain the binary hand image. The threshold value was computed automatically using the Otsu method. Bounding box operation is applied to the binary hand image to get the smallest rectangle which contains the binary hand image. The original hand image, binarized image, and the bounded image shown in Figure 1 (a), (b), and (c), respectively. The centroid of bounded image is computed and based on this centroid, the bounded binary and original images are segmented with 200 x 200 pixels. The segmented image and its centroid position are shown in Figure 1 (d) and (e). The centroid of the segmented binary image is computed and based on this centroid the ROI of palmprint can be extracted. The first and the second positions of centroid in binary and gray level image are shown in Figure 1 (f) and (g). This method is very simple. This method has been tested for 1050 palmprint images acquired from 210 persons, and the results show this that method is reliable.

Before the feature extraction phase, the extracted ROI are normalized using normalization method in [11] to reduce the possible imperfections in the image due to non-uniform illumination.

4. Feature Extraction

There are three main steps to extract the palmprint features based on fractal codes proposed in this paper. These steps are explained as follows.

4.1 Extraction of fractal codes of palmprint images

Fractal codes of palmprint images are obtained using the partitioned iterated function system (PIFS) method. In PIFS method, each image is partitioned into its range blocks and domain blocks. The size of the domain blocks is usually larger than the size of the range blocks. The relation between a pair of range block (R_i) and domain block (D_i) is noted as

$$R_i = w_i(D_i) \tag{1}$$

w_i is contracted mapping that describes the similarity relation between R_i and D_i , and is usually defined as an affine transformation as bellow:

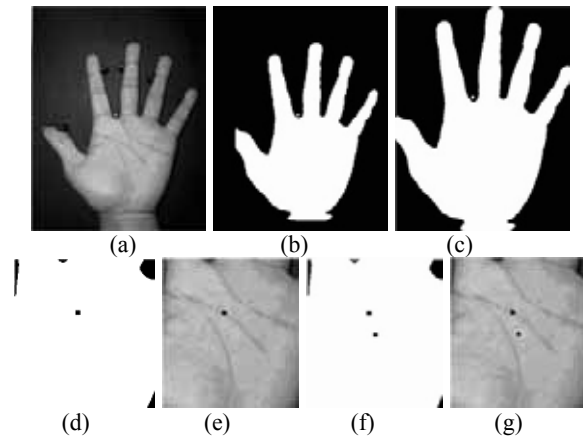


Figure 1. Extraction of palmprint, (a) original image, (b) binary image of (a), (c) object bounded, (d) and (e) position of the first centroid mass in segmented binary and gray level image, respectively, (f) and (g) position of the second centroid mass in segmented binary and gray level image, respectively.

$$\begin{matrix} x_i & a_i & b_i & 0 & x_i & e_i \\ w_i & y_i & = & c_i & d_i & 0 & y_i & + & f_i \\ & z_i & & 0 & 0 & s_i & z_i & & o_i \end{matrix} \tag{2}$$

where x_i and y_i represent top-left coordinate of the R_i , and z_i is the brightness value of its block. Matrix elements a_i, b_i, c_i and d_i , are the parameters of spatial rotations and flips of D_i , s_i is the contrast scaling and o_i is the luminance offset. Vector elements e_i and f_i are offset value of space. In this paper, we used the size of domain region twice the range size, so the values of a_i, b_i, c_i and d_i are 0.5. The actual fractal code p_i bellow is usually used in practice[19].

$$f_i = ((x_{D_i}, y_{D_i}), (x_{R_i}, y_{R_i}), size_i, \theta_i, s_i, o_i) \tag{3}$$

where (x_{R_i}, y_{R_i}) and (x_{D_i}, y_{D_i}) represent top-left coordinate position of the range block and domain block, respectively, and $size$ is the size of range block. The fractal codes of a palmprint image is denoted as follow:

$$F = \bigcup_{i=1}^N f_i \tag{4}$$

where N represents the number of the fractal code. The inequality expression bellow is used to indicate wether the range and the relevant domain block are similar or not.

$$d(R, D) \leq \epsilon, \tag{5}$$

where $d(R, D)$ represents rmse value, and ϵ is the threshold (tolerance) value. The range and the relevant domain block is similar if $d(R, D)$ is less or equal than ϵ . Otherwise, the block is regarded not similar.

4.2 Palmprint features representation

In this paper, we merely use the parameters (x_R, y_R) , (x_D, y_D) and the $size_i$ to form the palmprint features representation. The representation is formed using the rules bellow.

$$I(x_R, y_R) = \alpha_i, \text{ for } (x_R, y_R) \in F \text{ and } size_i = \min(size) \tag{6}$$

$$I(x_R, y_R) = 0, \text{ for } \dots \tag{7}$$

where I represents the palmprint features image and α represents the direction between the position of range and domain blocks that is computed as follows.

$$\alpha_i = \left| \frac{y_D - y_R}{x_D - x_R} \right| \tag{8}$$

- if $x_R < x_D$ and $y_R < y_D$ then $\alpha_i = \alpha_i$
- if $x_R > x_D$ and $y_R \geq y_D$ then $\alpha_i = 180 - \alpha_i$
- if $x_R > x_D$ and $y_R < y_D$ then $\alpha_i = 180 + \alpha_i$
- if $x_R < x_D$ dan $y_R > y_D$ then $\alpha_i = 360 - \alpha_i$
- if $x_R == x_D$ and $y_R < y_D$ then $\alpha_i = 90$
- if $x_R == x_D$ and $y_R > y_D$ then $\alpha_i = 270$

The criterion $size_i = \min(size)$ means the palmprint features representation is formed practically using the coordinate of the smallest size range block. Later, the representation is filtered as follow.

$$I'(x, y) = I(x, y) \cdot h(x, y)_{m \times n}, \tag{9}$$

$h(x, y)$ is filter which all of its component are one. Figure 2(b) show the palmprint features image of Figure 2(a).

4.3 Palmprint feature vector

Palmprint feature vector (V) is obtained by dividing the palmprint image into 16 x 16 blocks, and for each block its mean value is computed, so obtained the feature vector $V = (v_1, v_2, \dots, v_N)$, where $N = 256$, and v_i is mean value of block i .

Figure 2 (d) shows the palmprint feature representation in 16 x 16 sub blocks. Figure 3 shows three groups of palmprints from the same palm and palms with similar/different line structures. The features of these palmprints are plotted in figure 4. The results show that the features of three palm images from the same person are close to each other than the features of three palm images from the different persons with similar or different line structures.

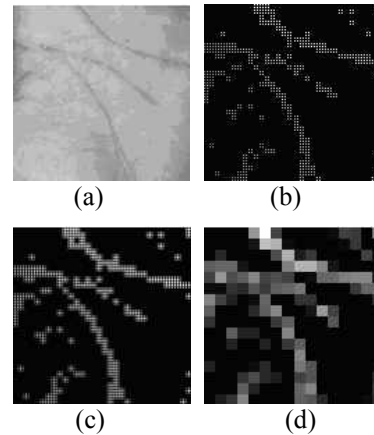


Figure 2. Palmprint feature extraction, (a) original image, (b) Image I, (c) Image I', (d) block feature representation

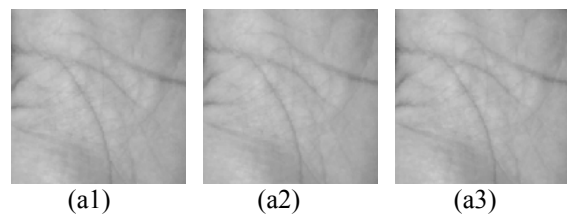
5. Palmprint Feature Matching

The degree of similarity between two palmprint features is computed as follows:

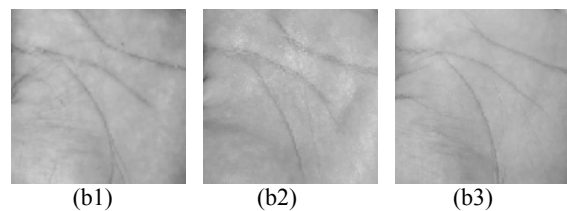
$$d_{rs} = 1 - \frac{(x_r - \bar{x}_r)(x_s - \bar{x}_s)}{[(x_r - \bar{x}_r)(x_r - \bar{x}_r)]^{1/2} [(x_s - \bar{x}_s)(x_s - \bar{x}_s)]^{1/2}} \tag{10}$$

where \bar{x}_r, \bar{x}_s are the mean of palmprint feature x_r and x_s , respectively. The above equation computes one minus normalized correlation between palmprint feature vector x_r and x_s . The values of d_{rs} are between 0 – 2. The d_{rs} will be close to 0 if x_r and x_s obtained from two image of the same palmprint. Otherwise, the d_{rs} will be far from 0.

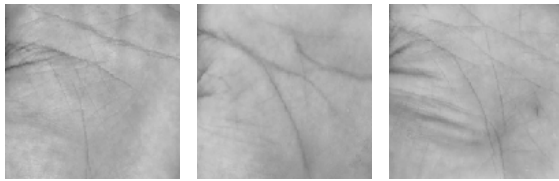
Figure 4 shows comparison of feature component of those palmprint shown in figure 3, and their score are listed in Table 1. The matching score of group A are close to 0, and the matching score of group B and C are far from 0. The average score of group A, B, and C are 0.1762, 0.5057, and 0.6452, respectively. It is easy to distinguish group A from group B and C using these scores.



Group A: palmprints from the same person



Group B: palmprints from different person with similar line structure



(c1) (c2) (c3)

Group 3: palmprints from different person with different line structure

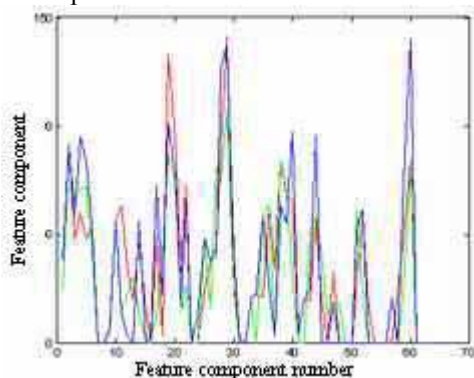
Table 1 Matching Score of groups A, B, and C in figure 3

	a1	a2	a3	Average
a1	0	0.1957	0.1404	0,1762
a2	0.1957	0	0.1925	
a3	0.1404	0.1925	0	
	b1	b2	b3	Average
b1	0	0.5352	0.3056	0,5057
b2	0.5352	0	0.6763	
b3	0.3056	0.6763	0	
	c1	c2	c3	Average
c1	0	0.6900	0.6177	0,6452
c2	0.6900	0	0.6280	
c3	0.6177	0.6280	0	

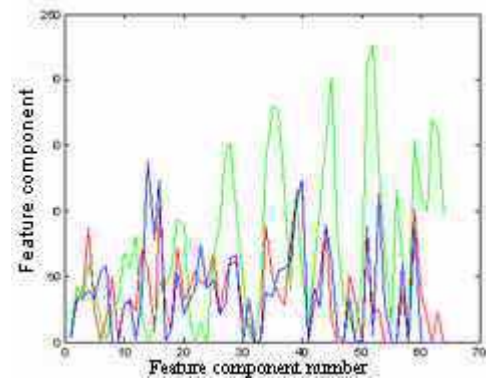
6. Experiments and Result

We collected palm image from 210 persons from both sexes and different ages, 5 samples from each person, so our database contains 1050 images. The resolution of hand image is 640 x 480 pixels. The palmprint images, of size 128 x 128 pixels, were automatically extracted from hand image as described in the Section 3. The averages of the first three images from each user were used for training and the rest were used for testing.

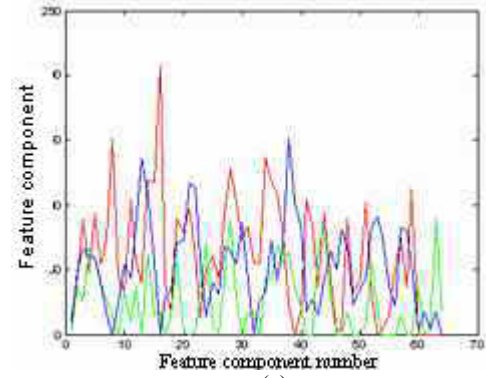
The performances of the verification system are obtained by matching each of testing palmprint images with all of the training palmprint images in the database. A matching is noted as a correct matching if the two palmprint images are from the same palm and as incorrect if otherwise.



(a)



(b)



(c)

Figure 4. Comparison of feature component of the palmprint group shown in figure 2. (a),(b),(c) are feature component of group A, B, and C, respectively. Red, green, blue color are the first, second, and third palmprint in each group, respectively.

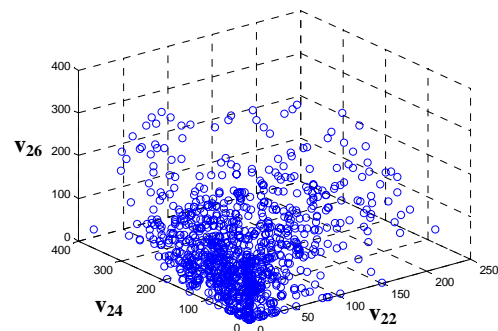


Figure 5. Distribution of three feature components of 1050 palmprints in feature space

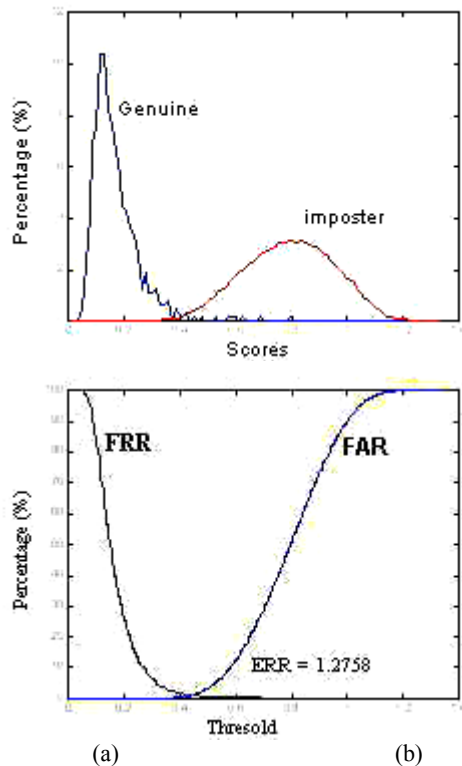


Figure 6. Performance of verification system, (a) genuine and imposter distribution, (b) FAR/FRR/EER with various threshold

Table 2. FRR/FAR with various threshold value

Threshold	FRR	FAR
0.4386	2.0734	0.4734
0.4586	1.9139	0.5158
0.4626	1.7544	0.6998
0.4746	1.4354	0.9160
0.4786	1.2759	1.3552
0.4986	1.1164	2.1480
0.5386	1.1164	2.2881

Figure 6 (a) shows the probability distributions of a genuine and imposter parts with tolerance value = 3, and feature vector length = 256 (16 x 16 blocks). The genuine and imposter parts are estimated from correct and incorrect matching scores, respectively. The result with various threshold and false acceptance rates (FAR)/false rejection rates (FRR) are shown in figure 6 (b). The equal error rate (EER) of the verification system is 1.2758. Table 2 show the performance (FAR/FRR) system with some threshold values.

7. Conclusions and Future Work

In this paper, we introduced a fractal characteristics based feature extraction and representation method for palmprint verification. The experiment results show that the proposed method can achieve an acceptable accuracy rate with FRR = 1.7544, and FAR = 06998 (see Table 2). In the

future, we will combine the proposed method with wavelet transformation to extract the feature of palmprint to retain the block operation.

8. References

- [1] Chih-Lung Lin., "Biometric Verification Using Palmprints and Vein-patterns of Palm-dorsum", <http://thesis.lib.ncu.edu.tw/etd-db/etd-search/>
- [2] Connie T., Andrew Teoh, Michael Goh, David Ngo, 2003, "Palmprint Recognition with PCA and ICA", http://sprg.massye.ac.nz/ivcnz/proceedings/ivcnz_41.pdf
- [3] Darma Putra, Adhi Susanto, Agus Harjoko, Thomas Sri Widodo, 2004, *Identifikasi Telapak Tangan dengan Memanfaatkan Alihragam Gelombang Singkat*, PAKAR, Vol.5, No.3, UTY, Jogjakarta.
- [4] Darma Putra, Adhi Susanto, Agus Harjoko, Thomas Sri Widodo, 2005, *Sistem Verifikasi dengan Biometrika Telapak Tangan*, Prosiding, UTY, Jogjakarta.
- [5] Darma Putra, Adhi Susanto, Agus Harjoko, Thomas Sri Widodo, 2006, *Biometrics Palmprint Verification Using Fractal Method*, EECIS proceedings, Part 2, pp.22-23, Brawijaya University, Malang, Indonesia.
- [6] Duta N., Jain A.K., Mardia K.V., 2002, *Matching of Palmprints*, Pattern Recognition Letters, 23, pp. 477-485.
- [7] Ekinci Murat, Vasif V., Nabiye, Yusuf Ozturk, 2003, *A Biometric Personal Verification Using Palmprint Structural Features and Classifications*, IJCI Proceedings of Intl, XII, Vol.1, No.1.
- [8] Jain A.K., 1995, *Fundamentals of Digital Image Processing*, Second Printing, Prentice-Hall, Inc.
- [9] Jain A.K., Ross A., and Pankanti S., 1999, *A Prototype Hand Geometry-based Verification System*, www.research.ibm.com/ecvg/publications.html
- [10] Jain A.K., *Introduction to Biometrics System*, <http://biometrics.cse.msu.edu/>.
- [11] Kumar A., David C.M.Wong, Helen C.Shen, Anil K.Jain, 2004, "Personal Verification using Palmprint and Hand Geometry Biometric", http://biometrics.cse.msu.edu/Kumar_AVBPA2003.pdf
- [12] LI Wen-xin, David Z., Shuo-qun XU., 2002, *Palmprint Recognition Based on Fourier Transform*, Journal of Software, Vol.13, No.5
- [13] Pang Y., Andrew T.B.J., David N.C.L., Hiew Fu San., 2003, *Palmprint Verification with Moments*, Journal of WSCG, Vol.12, No.1-3, ISSN 1213-6972, Science Press.
- [14] Sarraile, J., 2002, *Developing Algorithms For Measuring Fractal Dimension*, <http://ishi.csustan.edu>
- [15] Shu W., Zhang D., 1998, *Automated personal identification by palmprint*, Opt. eng., Vol. 37, No.8, pp. 2359-2363.
- [16] Tao Y., Thomas R.I., Yuan Y.T., *Extraction of Rotation Invariant Signature Based On Fractal Geometry*, <http://cs.tamu.edu>
- [17] Wohlberg B., Gerhanrd de Jager, 1999, *A Review of the Fractal Image Coding Literature*, IEE Transactions on Image Processing, Vol. 8, No.12.

- [18] WU Xiang-Quan, Kuan-Quan Wang, David Zhang, 2004, *An Approach to Line Feature Representation and Matching for Palmprint Recognition*, Journal of Software, Vol.15., No.6.
- [19] Yokoyama T., Sugawara K., Watanabe T., *Similarity-based image retrieval system using partitioned iterated function system codes*, The 8th International Symposium on Artificial Life and Robotics, January 24-26 2006, Oita, Japan, email:yokotaka@sd.is.uec.ac.jp
- [20] Yokoyama T., Watanabe T., Koga H., *Similarity-Based Retrieval Method for Fractal Coded Images in the Compressed Data Domain*, email:yokotaka@sd.is.uec.ac.jp
- [21] Zhang D., Wai-Kin Kong, Jane You, Michael Wong, 2003, *Online Palmprint Identification*, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.25, No.9.
- [22] Zhang D., and W.Shu, *Two novel characteristics in palmprint verification: datum point invariance and line feature matching*, pattern recognition vol 32, pp.691-702,1999

Synthetic Aperture Radar and Optical-Sensor Data Fusion Based On A Troust Wavelet Transform

Wihartini¹⁾ and Aniati Murni²⁾

- 1) Faculty of Engineering, Budi Luhur University
Jl. Raya Ciledug, Jakarta, Indonesia
Fax (62 21) 5853752, e-mail: wihartini@bl.ac.id
- 2) Faculty of Computer Science, University of Indonesia
Kampus UI, Depok 16424, Indonesia
Fax (62 21) 7863415, e-mail: aniati@cs.ui.ac.id

Abstract

This paper presents a continuous study on using complementary synthetic aperture radar (SAR) and optical-sensor data to obtain better information of land cover objects. The previous study show that the information fusion of multiresolution and multisensor segmentation based on the a trous wavelet transform gives a thematic image with artifacts. Data fusion based on multiresolution multispectral and panchromatic data using pyramidal wavelet transform gives good result, but the result is not so when it is applied to multisensor SAR and optical-sensor data. This paper proposed a methodology of data fusion using multiresolution and multisensor (SAR and optical-sensor) image based on a trous wavelet transform that includes noise elimination, a trous wavelet transformation, segmentation, high and low-frequency object extraction, data fusion and wavelet reconstruction. The result gives better land cover object information and discrimination.

Key words

Fusion., à trous wavelet transformation, segmentation, threshold.

1. Introduction

Synthetic Aperture Radar (SAR) is an imaging technique that employs an active and coherent microwave source of radiation to obtain target illumination [1,3,8]. SAR images are granular in appearance due to a phenomenon known as speckle noise. The speckle noise has reduced the ability of a human observer to resolve fine detail information within a SAR image. The use of SAR image to produce a thematic image becomes important due to a failure of using the optical-sensor images that contain cloud cover. Methodologies have been developed to use single sensor data or to fuse both sensor data to get better information [2,4,5,7,9].

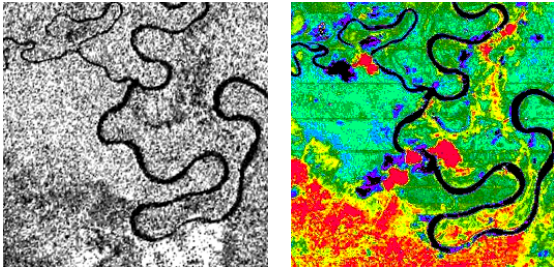
The a trous wavelet transform gives an offer to filter the low-pass and high-pass components of an image into different scales [6]. The wavelet-based and thresholding-technique segmented images are obtained in different scales. Different land cover objects may be represented well at different scales that make it possible to perform a data fusion of segmented images from different scales to produce a reasonable good thematic image [9]. Unfortunately, this approach produces a thematic image with artifacts.

Genderen and Pohl [5] have used a pyramidal wavelet transform and performed a data fusion of multispectral images and panchromatic image to get better land cover object boundaries. In this case, the

panchromatic image has finer spatial resolution compared to the multispectral images. The panchromatic information enhances the land cover object boundaries in the multispectral images. The process includes data fusion and reconstruction. This method has been adopted for fusing SAR and optical images which also have different spatial resolution sizes, but the result was not good. Inspired by [5] the methodology that used a trous wavelet transform [9] is improved by using the process of fusion and reconstruction applied in [5]. The process of this algorithm is non-pyramidal, and the transformation is stationary so it is not converging on a direction. Because the process is non-pyramidal, the image size remains unchanged during transformation process and the resulting image will be redundant on each scale.

2. Approach and Methods

The test data used for experiment is JERS-1 SAR L-band and SPOT panchromatic images of Central Kalimantan area at the longitude 113053' to 114000' and the latitude 2008' to 2011'. The SAR data is from path-row 96/304 and has been processed with a standard geocoded image resampled to UTM (Universal Transverse Mercator) projection. 25 ground control points have been used to obtain a registered images with an accuracy of less than 1 pixel root-mean-square error, and resampled with a spatial resolution of 12.5 meter.



a). SAR Jers-1 b) Landsat TM
Figure 1. Map of Central Kalimantan

The process consists of 5 steps: (i) speckle noise elimination using a trous algorithm and multiresolution support; (ii) a trous wavelet transform; (iii) wavelet-based segmentation; (iv) object segmentation using thresholding technique; (v) fusion of selected multiresolution and multisensor object segmented images and wavelet reconstruction.

3. À trous Wavelet Transform

3.1. À trous algorithm

A wavelet transformation for discrete data can be obtained by a version known as the à trous algorithm, which means by “hole”. The à trous algorithm is a non-dyadic resolution; it is a redundant or stationary transformation because there is no decimation process. Wavelet transformation with à trous 1-dimensional algorithm will produce one set $\{w_j\}$ on scale j , which contains the information detail and a scalar product c_j that contains image approximation information. Both w_j and c_j has the same pixel number with the original image.

$$c_0(k) = c_{np}(k) + \sum_{j=1}^{np} w_j(k) \quad (1)$$

For example, data $c_0(k)$ is a scalar product on pixel k from image function $f(x)$, with translated scaling function $\phi(x-k)$

$$c_0(k) = \langle f(x), \phi(x-k) \rangle \quad (2)$$

then the scalar product on scale j as follows :

$$c_j(k) = (1/2^j) \langle f(x), \phi(2^{-j}(x-k)) \rangle \quad (3)$$

if the wavelet function $\phi(x)$ follows the dilation equation $\frac{1}{2}\phi(x/2) = \sum_l g(l)\phi(x-l)$

$$\frac{1}{2}\phi(x/2) = \sum_n h(n)\phi(x-n) \quad (4)$$

Filter coefficient $h(k)$ can be obtained from the above dilation equation $\phi(x)$. From equations (2) and (3), the distance between samples increases with the factor of 2 from scale $(j-1)$ for $j>0$ to scale j . $c_j(k)$ can be calculated by iteration with the following relation :

$$c_{j+1}(k) = \sum_n h(n) c_j(k+2^j n) \quad (5)$$

where $h(n)$ is the spline function [3] (Strang and Nguyen, 1996) :

$$h(n) = \{1/16; 1/4; 3/8; 1/4; 1/16\}; -2 \leq n \leq 2 \quad (6)$$

and the wavelet coefficients are obtained from the difference of next two approximates is

$$w_j(k) = c_{j-1}(k) - c_j(k) \quad (7)$$

if the value $(k+2^j n)$ from $c_j(k+2^j n)$ from equation (6) is negative, then mirror rule is used:

$$c(k-n) = c(k+n) \quad (8)$$

while the image reconstruction can be done with the formula (2).

3.2. Denoising

Denoising or noise reduction process is different from filtering. Filtering is done in frequency domain, which is, disposing the unwanted frequency area by thresholding where some information is disposed. But denoising is a process to dispose noise by pyramidal algorithm (Moorlet) or non-pyramidal (à trous) algorithm.

The modeled image contains two components i.e. deterministic and stochastic, where the first is a signal while the second is a noise. There are two noise models: Gaussian additive noise model and non-Gaussian multiplicative noise model. Speckle is a non-Gaussian multiplicative noise. In SAR images, the noise is multiplicative, so components are multiplied. Homomorphic is applied to separate the deterministic and stochastic components, by taking the logarithm of the image:

$z = x \cdot v$ before processing, as follows:

$$\log(z) = \log(x) + \log(v) \quad (9)$$

where: $\log(z)$ is the new image, $\log(x)$ is the signal, and $\log(v)$ is the additive noise.

Since $\log(v)$ is an additive noise, then it can be treated in Gaussian model. Subsequently, for measuring the success level of speckle reducing process, it is done statistically. Generally, every image contains noise, so the wavelet coefficient also contains noise. The noise in the image follows the Gaussian or Poisson distribution, or the combination of both. In this study, multi-resolution support will be used to determine the noise significance.

The appropriate standard deviation value in scale j (σ_j) in the succession of wavelet planes is assessed from the standard deviation of the noise σ in the original image I , and from study of the noise in the wavelet space. This study consists of simulation an image containing Gaussian noise with a standard deviation equal to 1 and taking the wavelet transform of this image. Then, compute the standard deviation σ_j^e at each scale and get the curve of σ_j^e as a function of j will be give the characteristics of the noise in the wavelet space. The result from the wavelet transformation is:

$$\sigma_j = \sigma \sigma_j^e \quad (10)$$

if the distribution $w_j(x,y)$ is a Gaussian, with the mean = zero and standard deviation = σ , then it is possible to

compare $w_j(x,y)$ to $K\sigma$. Frequently, the value of K is set to 3, which is corresponding to $\epsilon = 0.002$. If $w_j(x,y)$ is small, it means insignificant and could be noise free. If $w_j(x,y)$ has a considerable value, it means significant.

If $|w_j(x,y)| \geq K\sigma$ then $w_j(x,y)$ is significant
 If $|w_j(x,y)| \leq K\sigma$ then $w_j(x,y)$ is not significant

$$(11)$$

So it is needed to estimate in the form of Gaussian noise model, and the standard deviation for each scale of multi-resolution support, which defined as follows:

$$M(j, x, y) = 1 ; \text{ if } w_j(x,y) \text{ significant} \\ = 0 ; \text{ if } w_j(x,y) \text{ not significant} \quad (12)$$

To determine the significance coefficient with $M(j,x,y)$, the masking is as follows :

$$\sum_{j=1}^{np} w_j(x, y) M_j(x, y) \quad (13)$$

and to reconstruct the image from the significance coefficient with the image c_{np} is expressed as:

$$c_o(x, y) = c_{np}(x, y) + \sum_{j=1}^{np} w_j(x, y) M_j(x, y) \quad (14)$$

The use of multi-resolution support, causes $c_0(x,y)$ seem to have residual artifact structure. To reduce residual artifact, used formula below

$$|\sigma_{R(n+1)} - \sigma_{R(n)}| / \sigma_{R(n+1)} < \epsilon \quad (15)$$

where:

$\sigma_{R(n+1)}$ is standard deviation of residu image R on $(n+1)$ iteration

ϵ is significant error (ϵ) which is less then 0.01

3.3 Image segmentation

The segmentation process is performed after à trous wavelet transformation is executed on both SAR and Landsat images. The transformation is done to observe the dispersion of objects in either Optical and SAR detail images in different scales. In transformed detail images w_3 and w_4 from Landsat, the objects are began to seem blurred while in SAR detail images w_1 and w_2 the objects were not yet clearly visible. Therefore, Landsat images $w_1, w_2,$ and $c_4,$ and SAR images w_3 and w_4 are selected for data fusion.

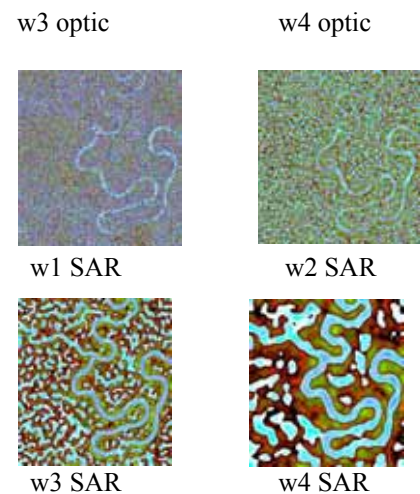
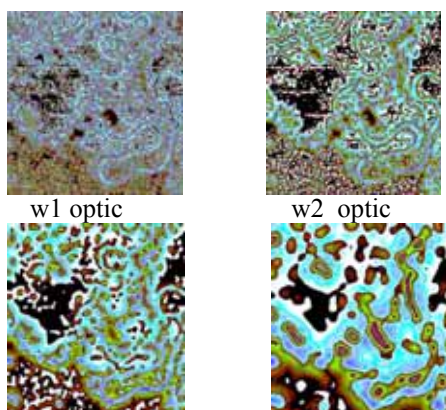


Figure 2. Detail images w_1, w_2, w_3 and w_4 from Landsat image and w_1, w_2, w_3, w_4 from SAR image

Subsequently, thresholding process is performed to select the objects to be filtered by observing the grey level of each segment of the detail images.

The image classification is done using the ERMAPPER software since ground truth is not available, thematic map and unsupervised classification is used with 7 predetermined classes (Table 1.) there are river(water), swamp, woodland, tropical grass, scattered tree, mangrove and rice paddy.. Variant distribution for example that shown on Table 1 had been chosen as minimum and maximum value in threshold process on detail images w_4 for each classes.

Table 1. Unsupervised classification w_4

Class	W4 sar	w4 landsat
1	0-5	0-5
2	9-16	9-26
3	19-26	28-52
4	28-35	53-107
5	37-57	206-238
6	229-247	239-248
7	248-255	249-255

4. Results

The next process after segmentation is to select detail images to be reconstructed. In this research, data fusion process is performed by reconstruction process with $w_1, w_2,$ and c_4 detail images from Landsat and w_3, w_4 images from SAR. The reconstruction equation is described as:

$$c_0 = w_1 (\text{Landsat}) + w_2 (\text{Landsat}) + w_3 (\text{SAR}) + w_4 (\text{SAR}) + c_4 (\text{Landsat})$$

these combination s can be varied by changing the detail or approximation images.

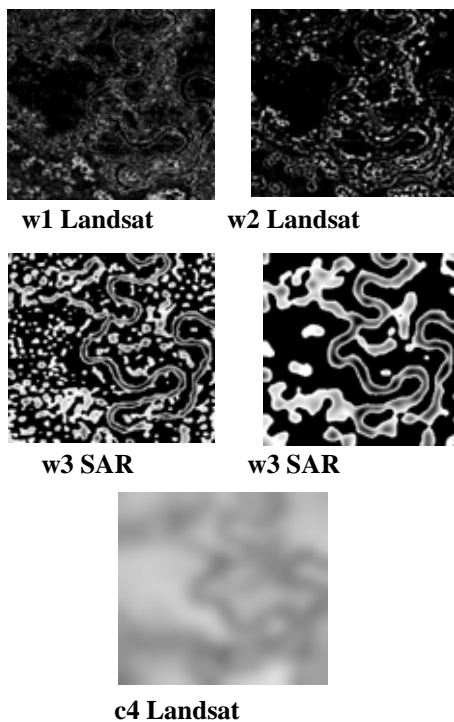
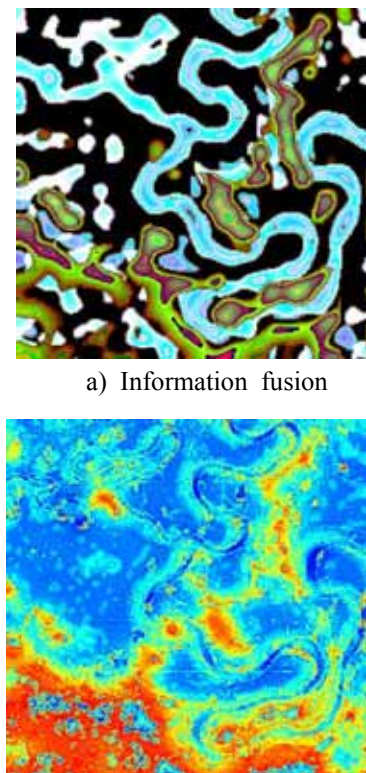


Figure 3 Data Fusion using à trous wavelet transform for the SAR and the optical-sensor images

Figure 4 below shows the difference between the result of data fusion and reconstruction with the result of information fusion.



b). Data fusion

Fig. 4. The result of data fusion and reconstruction and the result of information fusion of detail image w4 SAR to w3 Landsat

From the fusion data image above, it is apparent that many objects are not visible in Landsat original image. With data fusion from selected SAR image, many objects became visible in fused reconstructed image while in information fusion, the process is between one of SAR detail images and one of Landsat detail images.

5. Conclusion

- The use of multiresolution technique is very helpful to assist in observing objects in different scales.
- Thresholding technique is needed in data fusion process for reconstructed images to be as expected.
- The use of à trous wavelet transformation both in the denoising process and segmentation process can give several alternative segmented images.
- Data fusion the chosen two sensor segmented image can be done to obtain a reasonably better thematic image

References

- [1] Bruzzone, D.F. Prieto, and S.B. Serpico, "A Neural-Statistical Approach to Multitemporal and Multisource Remote Sensing Image Classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol 37, pp.1350-1358, May 1999.
- [2] A. Murni, A.K. Jain, and J. Rais, A Framework for Multidate Multisensor Image Interpretation, *Proceedings of International Geoscience and Remote Sensing Symposium (IGARSS'96)*, Lincoln, Nebraska, pp. 1851-1854, 1996.
- [3] C. Pohl and J.L. Van Genderen. *Multisensor Image Fusion in Remote Sensing: Concepts, Methods and Applications*. International Journal of Remote Sensing Vol 19. 1998.
- [4] S. Sardy, Wihartini, and S. Jatno, The Area Change Detection on Synthetic Aperture Radar Images by using Wavelet Transform, *Proceedings of the IASTED International Conference*, Marbella, Spain, September 4-7, 2001.
- [5] A.H. Schistad. Solberg, A.K. Jain, and T. Taxt, Multisource classification of remotely sensed data fusion of Landsat TM and SAR images, *IEEE Transactions on Geoscience and Remote Sensing*, 32(4):768-778, 1994.
- [6] Ulaby et al.
- [7] W. Sun, V.Heidt, P. Gong and G. Xu, Information Fusion for Rural Land-Use Classification With High-Resolution Satellite Imagery, *IEEE*

Transaction on geoscience and remote sensing,
Vol. 41 No. 4, pp. 883-890, April, 2003.

- [8] Wihartini and A. Murni, Informatiuon Fusion Based on Multiresolution and Multisensor SAR and Optical Sensor Segmented Images, *Proceedings of the Map Asia Conference*, 2005

A Study on Model Adaptation Technique for Improved Accuracy in Coded Speech

Nachiappan* and Abdul Manan Ahmad**

* Faculty of Computer Science and Information Systems, University Technology Malaysia, Skudai, Malaysia.
(Tel: +60(012)-6967860; Email: nachil82001@hotmail.com)

** Faculty of Computer Science and Information Systems, University Technology Malaysia, Skudai, Malaysia.
(Tel: +60(07)-5532070; Email: manan@fksm.utm.my)

Abstract

This paper looks into the possibility of using model adaptation technique to improve the recognition accuracy for coded speech. An unsupervised acoustic model using the Maximum Likelihood Linear Regression (MLLR) adaptation technique to improve the recognition performance of coded speech. In this paper, the effects of this LPC based speech coders in the performance of a speech recognition performance will be identified. It is also a known fact that the quality of speech will degrade when it is being encoded/decoded in several phases over the transmission channel. Distortion introduced by speech encoders can also be thought of as a mismatch between the training and testing conditions so we purpose to adapt the acoustic model and language model and produce a better results compare to the unadapted speech data. Purpose of this research is to study and examine the effects of LPC speech coders technique on speech recognition performance and how the MLLR adaptation technique could help in improving the recognition performance. There is two popular adaptation technique used to improve the recognition performance which is Maximum a Posteriori training (MAP) and Maximum Likelihood Linear Regression (MLLR) which will be reviewed in this paper and the reason for choosing MLLR technique will be explained. Two different speech coders from LPC codec family such as GSM full rate codec and CELP FS1016 codec will be used in this research to create the speech coded database which will be used in training and testing phase. Objective of this research is to improve the performance of a coded speech data by using unsupervised acoustic and language model adaptation technique. Toolkit used in this research is Hidden Markov Model Toolkit (HTK). The Hidden Markov Model Toolkit (HTK) is a portable toolkit for building and manipulating hidden Markov models. Feature extraction used at the speech recognition is based on the widely used Mel frequency cepstrum (MFCC). Baseline system will be the recognition performance of direct MFCC which is extracted from the coded data which uses acoustic model which is trained by the coded data. Comparison of recognition performance will be made between the results obtained from the baseline system and results after testing with the adapted acoustic and language model. This adaptation process is done to the coded data which later will be used to develop a single acoustic model for both type of coded data. DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus will be used in training and testing process. This research is being conducted to verify the hypotheses formed by reading previous research which suggest that recognition performance of coded speech can be improved significantly when adaptation technique used in training process and to prove that an adapted acoustic model can produce acceptable results for both type of coded data.

Keywords : Hidden Markov Model, acoustic model adaptation, language model adaptation, maximum likelihood linear, speech coding

1. Introduction

Cellular communications have been enjoying a tremendous worldwide growth and there is a great deal of R&D activity geared towards establishing global portable communications through wireless personal communication networks. Today there is a trend toward integrating voice-related applications (e.g. voicemail) on desktop and portable personal computers often in the context of multimedia communications. Most of these applications require that the speech signal is in digital format so that it can be processed, stored, or transmitted under software control. Digital speech signal which is uncompressed with a

high data rate requires high transmission bandwidth and storage. *Speech Coding* is the field concerned with obtaining compact digital representations of voice signals for the purpose of efficient transmission or storage. Speech coding involves sampling and amplitude quantization. The objective in speech coding is to represent speech with a minimum number of bits while maintaining its perceptual quality. Speech coding research started over fifty years ago with the pioneering work of Homer Dudley [1][2] of the Bell Telephone Laboratories. The motivation for speech coding research at that time was to develop systems for

transmission of speech over low-bandwidth telegraph cables.

One of the principal sources of degradation in the speech recognition applications deployed in mobile environments performance has degradation due to the presence of coding-decoding processes in the communication link. This performance degradation is due to the distortion introduced to the reconstructed speech signal by the quantization and bit rate reduction that takes place in the speech codec. It has been shown [3, 4] that, in general, the effect of coding on recognition accuracy tends to increase as the bit rate of the compressed signal decreases. As more speech applications are deployed into wireless, mobile, and cellular environments, the problem of distortion due to coding will become more widespread.

Cepstrum mean normalization (CMN) is known as an effective technique for removing convolutional linear channel distortion and is used as a standard technique in telephone speech recognition systems [5]. The problem is distortion caused by current low-bit rate speech LPC based codec such as GSM and CELP is nonlinear and the characteristics of each codec are different. Therefore, removing distortion by applying only a conventional CMN which assumes distortion to be stationary linear transfer on the cepstrum domain is difficult.

The purpose of this research is to focus on reducing the effect of this distortion made by the coded speech on recognition accuracy by making use of the MLLR and MAP adaptation technique when training the acoustic model. MLLR[6,7] is a very popular scheme and it has been widely used. MLLR can obtain a large improvement of the recognition rate over a speaker-independent model. The combination of MLLR and MAP is also proposed. In general, to obtain a high improvement, a lot of adaptation data with the phoneme transcription are needed and it takes time for adaptation. MLLR and MAP adaptation techniques which were originally developed for speaker adaptation, acoustic model for cellular phone speech can be trained from clean speech acoustic model with only a small amount of speech data targeted to a cellular phone network [8].

In this paper, Section 2 will discuss about the previous research in this area and identify research problem statement, Section 3 discuss about linear prediction-based coding of speech as well as some specifics about two low bit rate LPC based speech codec used in this research which is GSM (GSM 06.10) and CELP FS1016 while, Section 4 discusses about proposed model adaptation technique using MLLR and MAP. Section 5 explains about experiments in different condition and speech database that will be used to train and test the proposed technique. Finally Section 6 is summary of this paper.

2. Previous Research

Previous research conducted by other researchers shows that low bit rate compression of the speech waveform using the coders available today increases recognition errors [3,

4]. This problem can be solved using higher bit rate compression which less distorts the speech waveform but this will increase storage space which is limited on some mobile pervasive devices. Another problem is that existing recognition engine often has to be updated according to the speech compression employed by the digital communication network. Previous researchers suggest that retraining need to be performed using speech database processed by the coder whenever the speech compression algorithm changed [9].

Over the last 15 years, significant advances have been made in the area of speech coding and these speech coding algorithms are now able to produce communication quality speech at a bit rate as low as 2.4 kbits/s [10]. Since speech coders introduce distortion into the speech signal, it is only natural to expect that the recognition performance of these systems will deteriorate with the reduction in the bit rate [3]. In paper [11] the researcher concluded that robustness of speech recognition can be improved if one acoustic model trained for each speech codec. Paper [11] also proposed a new technique to reduce the mismatch between test and training data by training acoustic model with transcoded speech. This paper focus on how to solve the distortion created by low bit rate speech coding (GSM and CELP), and the effect of robust speech recognition to coding. A previous work [12], reported that an acoustic model which is adapted by coding speech could improve the recognition performance of coding speech.

Research question that may rise here is about degradation in performance that may happen because of speech coding process done to the speech database before recognition process. In this research LPC based speech coders like GSM full rate codec (GSM 6.10), CELP FS1016 codec (CELP 3.3) will be tested and our motif is to improve the performance of this both coded data and compare the performance among this coded data in matched and mismatched condition. Matched condition means each speech codec will have its own acoustic model and testing data will be same type of data. Results also will be collected for both type of coded data by using one uniform acoustic model which is trained and adapted using the proposed technique which is explained in section 4. Mismatched condition means testing coded data to general acoustic model which is trained using uncoded speech (direct MFCC feature vector).

3. Linear Prediction Based Coding

Linear-prediction based speech codecs assume that the speech signal is the output of an all-pole filter excited by an input signal (called the short-term residual, or excitation signal). Using transform notation we can use this model to express the resulting speech signal as the product of an excitation signal and a spectral shaping filter. To separate the spectral envelope of the speech signal from t

he excitation component, speech codecs first perform an LPC analysis on the speech signal and obtain the LPC polynomial. Then, the excitation component is obtained by inverse filtering the speech signal. The LPC filter is an all-pole filter whose coefficients can be found by minimizing the mean squared error between the predicted and true value of the original speech signal. Based on this observation, speech codecs approximate the residual in a given subframe according to its resemblance to the residual in previously reconstructed subframes [13].

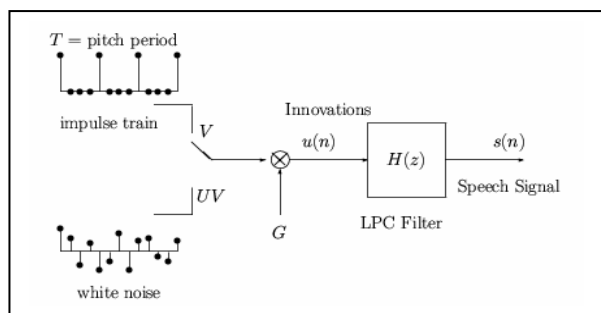


Fig 1. LPC Model For Speech Compression

3.1. The Full Rate GSM Speech Codec

Full-rate GSM speech codec [14] is a lossy speech coding decoding algorithm based on a regular pulse excited long term prediction scheme. The 8-kHz speech signals enter the codec where they are analyzed in frames of 160 samples from which 8th-order LPC parameters are obtained every 20 ms, producing an LPC analysis rate of 50 frames per second. Hence, the GSM coding algorithm produces a gross bit rate of 13.0 kbps, although the actual GSM transmitted bit rate is higher due to added error recovery and packet information.

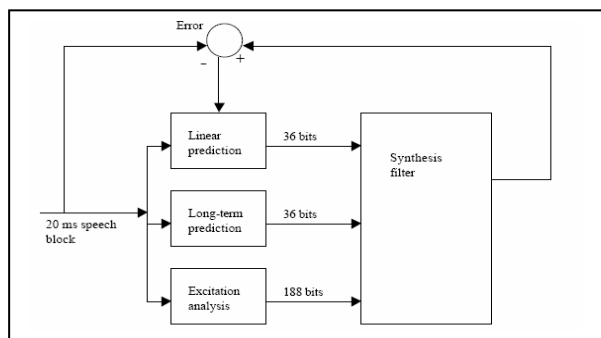


Fig 2. GSM Full Rate LPC-RPE codec

3.2. FS-1016 Celp Codec

The FS-1016 standard is a 4.8-kbps codec, based on a 10th order LPC analysis, followed by a CELP representation of the short-term residual signal [15]. The long-term periodicity is modeled by an adaptive codebook (which is an equivalent process to the long-term prediction block of the GSM codec). The adaptive codebook is generated from previous subframes of the reconstructed short-term residual. The long-term residual signal, or the difference between

the short-term residual and the short-term residual signal estimate, is coded by means of a fixed ternary stochastic codebook. This codec uses, as does the FR-GSM counterpart, 8 kHz as sampling rate. Its frame size is 30 ms long with four 7.5-ms subframes per frame.

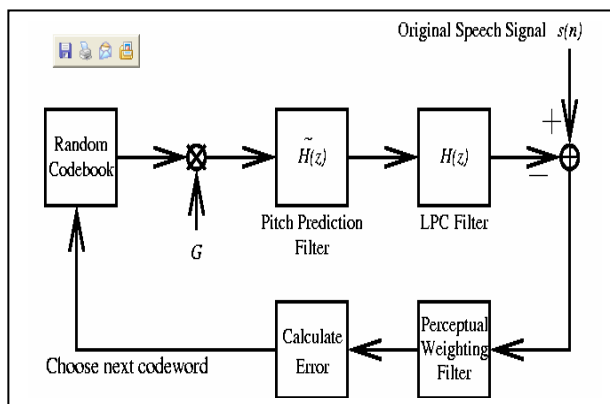


Fig 3. 4.8 kbps CELP Coder

4. Proposed Acoustic Model Adaptation

Adaptation schemes modify the parameters of the acoustic model directly in order to reduce the mismatch. Model adaptation to new speakers or environment conditions is commonly performed by the maximum a posteriori (MAP) [16] and the maximum likelihood linear regression (MLLR) [17] algorithms. In MAP adaptation, weighting is done between the old parameter set of the model and a maximum likelihood estimate based on the adaptation set. A key feature of MAP is that it only adapts models that are represented in the adaptation data. MLLR adaptation estimates a linear transform of model parameters to optimize the likelihood of the adaptation data. The method permits classes of models to share common transforms. Using classes enables adapting models that were not seen in the adaptation data. However, care must be taken to avoid clustering models that should not be adapted in a similar manner. The training material of an speaker independent automatic speech recognition (ASR) system is usually chosen to contain a wide range of different acoustic conditions such as :

- different speakers, speaking styles and accents,
- different transmission channels and room acoustics
- different types of microphones (close talking, headset)
- different levels and sources of ambient and channel noise

4.1. MLLR Adaptation

MLLR was originally developed for speaker adaptation [17, 18] but can equally be applied to situations of environmental mismatch. A set of transformation matrices are estimated which are applied to the Gaussian mean parameters. We have recently extended the approach so that the Gaussian variances can also be updated [19]. The

matrices are estimated so as to maximize the likelihood of the transformed models generating the adaptation data. The technique is implemented using the forward-backward algorithm and has close links with standard Baum-Welch training. The mean parameters are usually transformed by a full matrix (in the case of the HTK system a 40 39 matrix) or a block-diagonal matrix which accounts for only the correlations between the static's, 1st differentials and 2nd differentials as appropriate, while the variances are transformed either by a diagonal matrix (as in the experiments here) or by a more complex transform [19].

When only a small amount of data is available, or in cases where very robust transformation estimation is essential, each set of Gaussian parameters (means and variances) are transformed by a single matrix (single regression class case). As more data becomes available, or it is believed that the transcriptions are more reliable, more specific matrices can be computed using only the data that is aligned with that class. MLLR can be applied in a number of different modes including *unsupervised incremental* in which the system generates the labeling and updates the model parameters after every utterance (or after each small block of utterances) and *transcription mode* which processes complete sessions on block (static unsupervised adaptation) as used in this paper.

4.2. MAP Adaptation

Direct model adaptation techniques do not assume any underlying functional transformation but attempt to directly reestimate the model parameters. Acoustic units for which adaptation data is available are reestimated, leading to a *local* adaptation since unseen units are not modified. The most representative approach in direct adaptation is Bayesian learning, often implemented via maximum *a posteriori* (MAP) estimation, developed at Bell Labs [20,21]. MAP adaptation combines under a well defined mathematical formulation the information provided by the adaptation data with some prior knowledge about the model parameters described by a prior distribution. A fundamental property of MAP adaptation is its nice asymptotic convergence to maximum likelihood estimation when the amount of adaptation data increases. This convergence is however fairly slow and a large amount of adaptation data is needed to observe a significant performance improvement.

MAP estimation requires the definition of a prior distribution. It is convenient if the prior density is from the same family as the posterior distribution (the conjugate prior) if it exists. For mixture Gaussian HMMs such a conjugate prior of finite dimension does not exist and an alternative approach presented in [22] is usually used. One key advantage of the MAP approach is that as the amount of training data increases towards infinity the MAP estimate converges to the ML estimate. Its main drawback is that it is a *local* approach to updating the parameters i.e. only parameters that are observed in the adaptation data will be altered from the prior value.

4.3. Proposed Adaptation Technique

It is necessary to have sufficient data points to robustly estimate MLLR or similar transforms. If appropriate thresholds/forms of transforms are not used then poor performance (even poorer than speaker independent performance) can be the result due to over-training on the adaptation data. Of course, in practice, limits on the form of the transforms combined with a range of thresholds are used to ensure that performance is never poorer than SI. However it would be preferable to instead use schemes which don't, for instance, require switching from diagonal transforms to block-diagonal as more data becomes available. Various solutions to this problem have been suggested and all increase the applicability for MLLR for rapid adaptation. The solutions include a somewhat ad-hoc MAP-like interpolation between the original mean and the MLLR estimated mean [23].

In both [24] and [25], it is suggested that a prior distribution for the mean transformation matrix parameters be used (dubbing the technique MAPLR) and this improves performance when very small amounts of data are available. The prior distribution can be estimated by generating transforms from the set of training speakers. The above approach use a MAP style estimation approach for MLLR parameters.

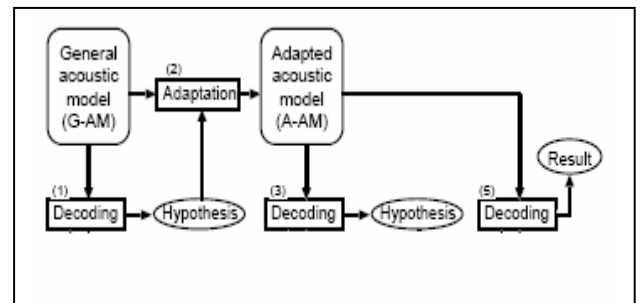


Fig. 4. An overview of the recognition flow using both acoustic model adaptation.

Traditionally, transformation-based model adaptation techniques like MLLR are based on a maximum likelihood formulation. It is however reasonable to assume that some values of the transformation parameters are more likely than others, so that by introducing additional information about the transformation parameters, improved performance can be expected. This additional information can take the form of a prior distribution of the transformation parameters, which is then used to constraint the estimation via the use of a MAP estimation criterion, rather than maximum likelihood. We will reformulate the traditional MLLR algorithm under such a Bayesian framework by introducing a prior distribution for the affine transformation matrices, leading to the MAPLR algorithm. The proposed acoustic model adaptation method was done using MAP and MLLR adaptation technique. Figure 4 shows the overview of the

recognition flow which includes acoustic model adaptation processes. The overall steps are as follows:

1. Recognizing all utterances using a general speaker independent acoustic model (G-AM).
2. Building a speaker adapted acoustic model (A-AM) by adapting the G-AM by the MLLR and MAP adaptation technique using the recognition results obtained in 1.
3. Obtaining improved recognition hypothesis by re-recognizing the utterances using A-AM.

5. Proposed Experiment Condition

A way to improve the performance of GSM and CELP coded data is to adapt the existing LPC based models to the decoded data from the used codecs. We investigate the effect of MLLR+MAP adaptation technique on the speaker recognition performance on 2 different conditions:

- Condition A : Mixed adaptation, where the adaptation data are samples decoded from different codecs such as GSM and CELP.
- Condition B : Mismatched condition where the trained data is direct MFCC feature vector and tested data is coded speech such as GSM and CELP.

To evaluate the proposal method, we compared recognition performance of the following HMM :

- baseline HMM,
- the coded speech HMM using proposed adaptation technique

Baseline HMM is trained by uncoded speech. The coded speech HMM is trained with coded training data. This is the upper limit of the proposal method. Baseline acoustic model is trained using Wall Street Journal speech database of phonetically balanced sentences. General acoustic model is trained for each speech coded data using DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus. First this speech data will be coded using GSM and CELP and then it will be used to train acoustic model for each speech codec. Accuracy is calculated for matched and mismatched condition and comparison is done to identify whether the proposed adaptation technique improved speech recognition performance for coded speech. The recognition performance is evaluated by word accuracy using equation below, where N is the total number of words, D is the number of deletions, S is the number of substitutions, and I is the number of insertions.

$$Accuracy = \frac{N - D - S - I}{N} \cdot 100 (\%) \quad (1)$$

6. Summary

This paper gives basic information about LPC based speech

codec such as GSM 06.10 and FS-1016 CELP which is used to encoding and decoding the speech. This paper also identifies some of the unresolved problem by previous researchers. Main problem identified is the distortion introduced by different low bit rate speech codec which degrade the recognition performance. Another problem identified is mismatched condition between test and train data introduce degradation in performance. Some researchers proposed that each speech codec must have a specific acoustic model in order to produce high recognition rate. But this paper try to solve the identified problem by proposing one acoustic model which is adapted using the proposed technique for both speech codec. The chosen adaptation technique is MLLR+MAP which used to create an identical acoustic model for both speech codec. Comparison is will be done between the matched and mismatched condition to find whether the proposed method really improved speech recognition performance. Some of the popular adaptation technique such as MLLR and MAP is discussed in this paper and combination of both technique is proposed to design the acoustic model.

7. References

- [1] H. Dudley, "Remaking Speech," J. Acoust. Soc. Am., Vol. 11, p. 169.
- [2] H. Dudley, "The Vocoder," Bell Labs. Record., 17, p. 122, 1939.
- [3] Euler S., Zinke J., "The Influence of Speech Coding Algorithms on Automatic Speech Recognition", IEEE Transactions on Acoustics, Speech and Signal Processing. Vol ASSP- No. 1994.
- [4] Lilly B. T., Paliwal K. K., "Effect of Speech Coders on Speech Recognition Performance", Proc. ICSLP-96, 1996.
- [5] A. Acero, Acoustical and environmental robustness in automatic speech recognition," Kluwer Academic Publishers, Boston, 1993.
- [6] C.J. Leggetter and C. Woodland. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. Computer Speech and Language, Vol. 9, pp. 171-185, 1995.
- [7] M. Padmanabhan, L.R. Bahal, D. Nahamoo, and M.A. Picheny. Speaker Clustering a Transformation for Speaker Adaptation In Large-Vocabulary Speech Recognition System. In Proceedings of the ICASSP, pp. 701-704, 1995.
- [8] R. Gruhn, H. Singer, H. Tsukada, A. Nakamura, M. Naito, A. Nishino, Y. Sagisaka, S. Nakamura, Cellular Phone Based Speech- To-Speech Translation

- System ATR-MATRIX," Proc. of ICSLP 2000, Vol. IV, pp. 448-451, 2000.
- [9] Dan Chazan, Gilad Cohen, Ron Hoory and Meir Zibulski, "Low Bit Rate Speech Compression for Playback in Speech Recognition Systems," *European Signal Processing Conf. (EUSIPCO)*, pp. 1281-1284, 2000.
- [10] W.B. Kleijn and K.K. Paliwal (eds.), *Speech Coding and Synthesis*, Elsevier Science B.V., Amsterdam, 1995.
- [11] L. Besacier, C. Bergamini, D. Vaufreydaz, E. Castelli "The effect of speech and audio compression on speech recognition performance" in *IEEE Multimedia Signal Processing Workshop*, Cannes, France, October 2001.
- [12] Tanaka, K., Kuroiwa, S., Tsuge, S. and Ren, F. 2002, "The Influence of Speech Coders for IP Telephone on Speech Recognition" *Information-2002*
- [13] Huerta, Juan M. / Stern, Richard M. (2000): "Instantaneous-distortion based weighted acoustic modeling for robust recognition of coded speech", In *ICSLP-2000*, vol.3, 842-845.
- [14] European Telecommunication Standards Institute, "European digital telecommunications system (Phase 2); Full rate speech processing functions (GSM 06.01)", ETSI 1994
- [15] Campbell Jr. J.P., Tremain T.E., Vanoy C. Welch, "The DoD 4.8 KBPS Standard (Proposed Federal Standard 1016)", in *Advances in Speech Coding* edited by Atal B. Cuperman V. and Gersho A. Kluwer Academic Publishers 1989.
- [16] Gauvain J., "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 2, pp. 291-298, April 1994.
- [17] Leggetter C. and Woodland P., "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models," *Computer Speech and Language*, vol. 9, no. 2, pp. 171-185, April 1995.
- [18] Leggetter C.J. & Woodland P.C. (1995). Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models. *Computer Speech & Language*, Vol. 9, pp. 171-185.
- [19] Leggetter C.J. & Woodland P.C. (1995). Flexible Speaker Adaptation Using Maximum Likelihood Linear Regression. *Proc. ARPA Spoken Language Technology Workshop*, pp. 104-109. Morgan Kaufmann.
- [20] Gales M.J.F., Pye D. & Woodland P.C. (1996). Variance Compensation Within the MLLR Framework for Robust Speech Recognition and Speaker Adaptation. In *Proc. ICSLP'96*. Philadelphia.
- [21] H. Lee, C.-H. Lin, and B. H. Juang. A study on speaker adaptation of continuous density HMM parameters. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pages 145-148, Albuquerque, New Mexico, April 1990. ICASSP'90.
- [22] L. Gauvain and C.-H. Lee. Maximum a posteriori estimation for multivariate gaussian mixture observations of Markov chains. *IEEE Transactions on Speech and Audio Processing*, 2(2):291-298, April 1994.
- [23] J.L. Gauvain & C.H. Lee (1994) Maximum A posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains. *IEEE Trans. SAP*, Vol. 2, pp. 291-298.
- [24] S. Goronzy & R. Kompe (1999) A MAP-Like Weighting Scheme for MLLR Speaker Adaptation. *Proc. Eurospeech'99*, pp. 5-8, Budapest
- [25] C. Chesta, O. Siohan & C.H. Lee (1999). Maximum A Posteriori Linear Regression for Hidden Markov Model Adaptation. *Proc. Eurospeech'99*, pp. 211-214, Budapest.

Minimizing Handoff Latency in AMTree Protocol

Ali M. Al-Sharafi¹ Mohamad Othman² Shamala Subramaniam²

Faculty of Computer Science and Information Technology
Universiti Putra Malaysia, Serdang 43400, Selangor D.E, Malaysia
Tel: +603-89466565, Fax :+ 603-89466577 / +603-89466572
¹a_alsharafi@hotmail.com, ²(mothman,shamala@fsktm.upm.edu.my)

Abstract

In active networks, programs can be injected to network elements (routers/switches). This allows programmers to enhance existing protocols or deploy new protocols. AMTree protocol is an active network based protocol that makes sending packets to receivers - after source migration - an efficient process. In AMTree protocol, after source migration, handoff latency computed from the time of reconnecting to a new base station until finding nearest core to the mobile source. In this paper we present two new algorithms to minimize the handoff latency in AMTree protocol. We show that handoff latency is much lower than that of AMTree handoff algorithm if the mobile source connects to a base station which is subscribed to the multicast group. Also we show that our algorithm gets better results than AMTree if the new base station is not in the multicast group.

Keywords:

Active networks, Mobile networks, Multicast, Handoff.

1. Introduction

Multicast is communication between a single sender (or multiple senders) and multiple receivers on a network. Recently, more and more group communication applications (e.g., video-conferencing, online-gaming, and long-distance education) have emerged with the increasing popularity of the Internet. To support such multi-user applications, multicast is considered as a very efficient mechanism since it uses some delivery structures (e.g., trees or meshes) to forward data from senders to receivers, aiming to reduce duplicate packets.

In traditional networks, the design of multicast protocols such as DVMRP [1], MOSPF [2], CBT [3] and PIM [4] was for fixed hosts in mind hence, it faces some problems in mobile networks. Active Networks [5] are a new paradigm for solving network problems. This paradigm uses the computational power at intermediate network nodes to facilitate processing of traffic passing through.

AMTree protocol [6] is an active network based approach to solve the problem of having mobile source in the multicast group. The algorithm to compute handoff latency in AMTree results in low latency compared with that of bidirectional home agent method [7].

In this paper we propose a new algorithm to reduce the handoff latency in AMTree.

In the next section we present a background about active networks and AMTree protocol. Section 3 describes the proposed method and algorithm. In section 4 we present the simulation model and the results. The last section concludes this paper.

2. Background

The idea of AMTree protocol is to use capabilities of Active networks to give the source of multicast group the ability of being mobile and to continue sending data packets to the tree after migration. This capability is executing user instructions in the network elements. This section presents a brief look at Active Networks and AMTree Protocol.

2.1 Active Networks

Active network is a new paradigm to network architecture in which the switches of the network perform customized computations on the messages flowing through them.

There are two possible approaches to build active networks. Programmable switches approach and capsule approach.

- Programmable Switches Approach. Here programs are injected into the programmable active node separately from the actual data packets that traverses through the network. 'User' would send the program to the network node

(switch or router), where it would be stored and later executed when the data arrives at the node, processing that data. The data can have some information that would let the node decide how to handle it or what program to execute [8, 9].

- Capsule Approach: in this approach, the program is integrated into every packet of data sent to the network. Each message or capsule contains a program fragment that may or may not have some embedded data. When these capsules arrive at the active node, it interprets the programs and sends the embedded data depending on its interpretation of these programs [10]. This concept is similar to Postscript code, where actual data is embedded in program fragments that the printer understands. In this approach, each active node would have built-in a mechanism to load the encapsulated code, an execution environment to execute the code and a relatively permanent storage where capsules would retrieve or store information.

2.2 AMTree Protocol

AMTree protocol is designed to solve some of the problems that may occur when there is a mobile source in the multicast group. AMTree takes advantage of the processing capabilities at routers which enable mobile source to continue sending packets to receivers after migration. This means that multicast tree can be maintained without much modification and incurs minimal packet latency. Hence, handoff latencies will be low and multicast tree is updated dynamically and efficiently.

After mobile source migration, it sends a message to old contact point (its local router before migration) to give its new address. The old contact point then multicast a message to the tree with the new address. Each core (core refers to any router with more than one subscriber) in the tree -after receiving the message- sends a CORE_CONNECT message to the new contact point as a designated core. This process uses the computation in each router to determine the nearest core to the mobile source's new address. The designated core is then used by the mobile source to unicast packets, the core in turn will multicast packets along the tree.

3. Proposed method

We have developed two algorithms to minimize the latency of handoff procedure in AMTree protocol. The two new algorithms take advantage of using computation capabilities of active networks. In order to minimize the number of messages sent using the algorithm proposed in AMTree protocol, we modified the existing acknowledgement message by adding the IP address of the first core in the path from the receiver to the source (which considered the nearest core to the receiver). Figure 1 shows the first algorithm.

Algorithm: (determine the nearest core when a base station subscribed to the group)

```

If a node is not a core and becomes a core
then
  Send N_CORE_MSG down to all
  subscribers
Else
  If a node is a core and received
  JOIN_MSG then
    Send ACK_MSG down to the receiver
  If a core node received a N_CORE_MSG
  then
    {
      Save the IP address which included in
      the message
      Discard the message
    }
  
```

Figure 1 Algorithm to determine the nearest core for each node subscribed to multicast group.

The algorithm works as follows:

When a receiver wants to join the multicast group, it indicates its interest in joining by sending a JOIN_MSG to its local router. The router will forward the message to the source through the shortest path. If a core router is found in the path it will send an ACK_MSG (containing its IP address as nearest core) down to the receiver. The receiver and all nodes (routers) in the path will save the core's IP address as their nearest core. When the mobile source migrate and reconnect to the new base station, it will get the IP address of the nearest core to this node (i.e. base station). The nearest core will be the core that receive the packets from the source as unicast and its job is to multicast the packets along the tree.. If the base station was not subscribed in the previous tree then it will send a message to all nodes connected to it to ask for the nearest core.

If a router but not a core becomes a core (new node has subscribed to it), then it sends an N_CORE_MSG to all nodes subscribed to it containing its IP address as nearest core

If a core router receives an N_CORE_MSG then it saves the IP address which is included in the message and discard the message. The IP address will be used when this core becomes no core any more (when it has only one subscriber left and other subscribers leaves the multicast group).

The above algorithm was proposed to be used when the mobile sources connected to a base station node that is subscribed to the multicast group.

The second algorithm was proposed to be used when the mobile source connected to a base station node that is

not subscribed to the group. Figure 2 shows the second algorithm.

Algorithm: (determine the nearest core when a base station is not subscribed to the group)

If a base station is not subscribed to the multicast group then
 Send GET_NEAREST_MSG to all nodes connected to it.

Each node subscribed to the tree reply to the message with its nearest core

If more than one reply received then
 Compare No. of Hops and choose the shortest one.

Figure 2. Algorithm to determine the nearest core when the base station is not subscribed to the tree.

The idea of this algorithm is to send a message from the base station to all nodes connected to it asking for the nearest core to each one. Each node that is subscribed to the multicast group will send the IP address of its nearest core. By comparing the number of hops the base station node choose the nearest core node which the mobile source unicast its packets to, then it multicast them to the tree.

If the base station is not subscribed to the multicast group, we still have high probability of having one of the nodes that is connected to it is subscribed to the group. Figure 3 shows the number of nodes subscribed to the multicast group versus number of receivers.

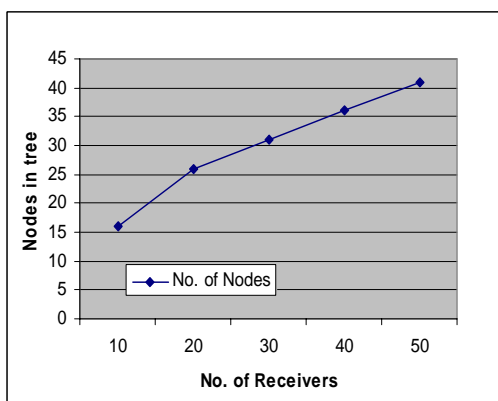


Figure 3. Number of nodes subscribed to the tree versus number of receivers.

4. Simulation and Results

We have tested our algorithm after re-simulating AMTree protocol. We have tested AMTree protocol and new algorithm on a network of 50 node mesh topology

with degree ranging from one to six with average three. The data rate of interconnecting links between active routers is set to 10 Mb/s. The wireless links have a data rate of 2 Mb/s. Bit errors at the wireless link are assumed to be handled by a data link layer protocol. Each base station is assumed to manage a cell and overlapping of cells means there are no silent areas. The mobile source controls the time of handoff, meaning the time of handoff and the duration of the handoff procedure can be measured. The mobile source migrates randomly to any of the receiver's subnet. Hence, the migrations are not necessarily local. There is only one ongoing session in any moment and the packets generated by the source are of size 1024 bytes at rate one packet per second.

The performance studied was on varying number of receivers, ranging from 10 to 50 and ten simulation runs were done and results averaged.

We compared the handoff latency achieved by AMTree protocol and after we applied our algorithms and we found out that our algorithms achieved much lower latency, either in case of a base station is subscribed to the multicast group or not subscribed. Figure 4 shows results of our first algorithm compared to the previous algorithm proposed in AMTree protocol. Figure 5 shows the results for both cases of the base station (i.e. subscribed or not subscribed).

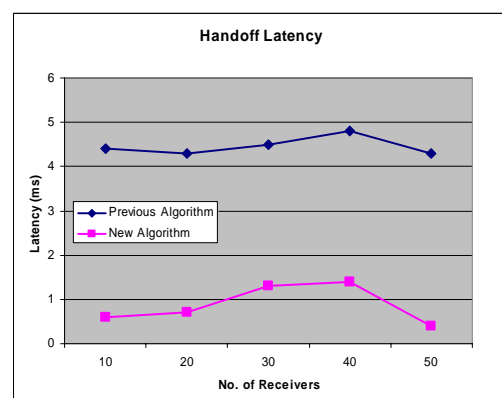


Figure 4. Handoff latency for varying receiver size.

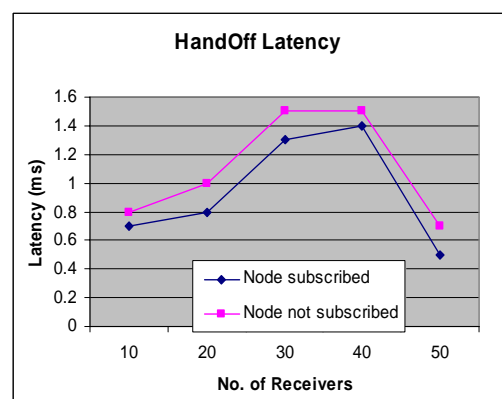


Figure 5. Handoff latency when node subscribed to the group and when it is not subscribed.

We manage to get almost the same results for other parameters after this improvement. Figure 3 shows an example of one of the parameters which is number of cores.

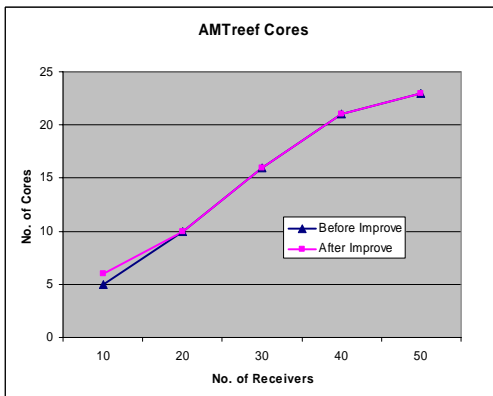


Figure 6. Number of cores vs. number of receivers.

5. Conclusion

In this paper we presented two new algorithms to minimize the latency for handoff procedure in AMTree protocol either the mobile source connected to a base station that is subscribed to the multicast group, or a base station that is not subscribed. We showed that the handoff latency achieved by our algorithms in both cases is much lower than that of previous algorithm proposed in AMTree protocol.

6. References

- [1] Waitzman, D., Patridge, C. and Deering, S. November 1988. Distance vector multicast routing. RFC1075.
- [2] Moy, J. 1994. Extension to OSPF. Internet draft 1584.
- [3] Ballardie, A., Francis, P. and Crowcroft, J., 1993. Core Based Trees (CBT): An architecture for scalable inter-domain multicast routing. In *SIGCOMM'93*, 85–95. San Francisco, CA.
- [4] Deering, S., Estrin, D., Farinacci, D., Jacobson, V., Liu, C.-G. and Wei, L. August 1994. An architecture for wide-area multicast routing. In *SIGCOMM'94*, 126–135.
- [5] Tennenhouse, D.L., and Wetherall, D.J., April, 1996. Towards an active network architecture. *Computer Communication Review* 26, 5–18.
- [6] Chin, K-W. and Kumar, M. 2001. AMTree: An active approach to multicasting in mobile networks. *Mobile Networks and Applications*, 6(4): 361–376.
- [7] Perkins, C. October, 1996. IP mobility support. RFC2002
- [8] Alexander, D.S., Shaw, M., Nettles, S.M. and Smith, J.M. 1997. Active bridging. In *Proceedings of SIGCOMM '97*.
- [9] Bhattacharjee, S., Calvert, K. L. and Zegura, E. April, 1997. An architecture for active networking. *High Performance Networking (HPN'97)*.
- [10] Wetherall, D.J., Gutttag, J. and Tennenhouse, D.L. April, 1998. ANTS: A toolkit for building and dynamically deploying network protocols. *IEEE OPENARCH'98*, San Francisco, CA.

A Study on Model Compensation Using Parallel Model Combination and Adaptation Technique For Improved Accuracy in Noisy Speech Recognition

Tan Sang Sang¹, Abd. Manan Ahmad²

¹Faculty of Computer Science & Information Systems
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5532359, E-mail: ss_hello33@yahoo.com

²Faculty of Computer Science & Information Systems
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5532359, E-mail: manan@fksm.utm.my

Abstract

Due to the mismatch between training and operating conditions, speech recognition systems often exhibit drastic performance degradation when they are practically used in real-world environments. Various techniques based on noise resistance, speech enhancement and model compensation approaches have been widely used to improve the performance of speech recognition in noise. Parallel Model Combination (PMC) is one of the most popular techniques, which has been proved powerful in compensating recognition models so that they reflect speech in noisy acoustic environments. However, studies have shown that some assumptions and approximations made in the PMC, primarily in the domain transformation and parameter combination processes, are not accurate in certain situations, thus restricting better performance. In this paper, we review and compare several adaptation techniques that possess the potential to be used along with PMC to achieve better accuracy. Among the various adaptation techniques, Maximum Likelihood Spectral Transformation (MLST) has been found superior to other adaptation techniques like MLLR and MAP. MLST gets the run upon because it requires only a small amount of adaptation data and the adaptation process is computationally inexpensive.

Keywords:

Parallel Model Combination (PMC), Maximum Likelihood Spectral Transformation (MLST), model compensation, adaptation technique, noisy speech recognition

1. Introduction

Nowadays, the performance of those state-of-the-art speech recognition systems has been significantly increased. Speech can be recognized with accuracy higher than 90% using the state-of-the-art speech recognition systems [1] [2]. Anyway, this is often the result gained in controlled environments, which are usually quiet environments. In real world application, it is rarely quiet and there is also little control over the acoustic environment. This is especially true in these days since speech recognition has been widely used on handheld devices like PDA and cellular phone. The portability of these devices means the users use speech recognition not only in the quiet office, but also in shopping malls, in cars, on the streets, etc. According to some estimates, there is a 20% - 50% decrease of recognition rates when speech recognition is implemented in a natural field environment [3] [4].

The principal cause of drastic performance degradation is the distinct difference between training and operating condition. This environmental difference is caused by noises, which are the sounds that are transmitted to a recognition system but are not part of the meaningful input signal. The predominant kind of noise is the background noise, which is produced at the speaking environment and being input to the recognition system along with the speech. Background noise is superimposed upon the speech input, i.e. it is additive. When speech signals are transformed and transmitted through some speech channels, channel noise is also induced on the spoken input. Two primary channels in speech recognition are microphones that are used during recording and telephone channels that serve as the communication channels. Environmental noises are different in their characteristics. Basically, signal-to-noise ratio (SNR) becomes an objective measure to estimate the amount of noise present with a speech signal. Studies show that performance of a recognizer is rather uniform for SNRs greater than 25dB but there is a very steep degradation as

the noise level increases. In training environment or laboratory, SNRs may reach 90dB (but 50dB would be more typical). Unfortunately, most real-world environments have poor SNRs. For instance, in factory or subway, the SNRs are usually less than 5dB [5]. This significant gap causes most recognizers to perform poorly in operating condition. Besides of SNRs which greatly related to the loudness of the noise, the accuracy of speech recognition also depends on the nature of the noise. For instance, non-speech noise is easier to deal with, compared to background conversation or interfering speech that make discrimination difficult. The variability of the noises, or how much the noises change, is another differentiating characteristic of noises. Steady or periodic noises (engine noise) are easier to identify and eliminated compared to impulsive noises (e.g. machine gun) because of their consistency [6]. In the case of channel distortion, characteristics of noises may also depend on the types of speech channel and the quality of the channel.

Besides of environmental disturbances, the speakers also introduce some difficulties to the task of speech recognition. During recognition, speech noise (e.g. lip smacks) and non-communicative sounds (e.g. speaker's hesitation, "Uh...") are produced by speakers. Another issue is the Lombard effect which describes the changes that occur to speech when speakers attempt to make themselves heard over noise. Some of the speech and acoustic features affected by a speaker's attempt to overcome the noise include increased vocal effort, greater duration of words due to increased vowel length, etc. The Lombard effect makes speech easier to hear and understand when the listener is another human, but in speech recognition, it has a negative impact that may decrease the accuracy up to 25% [5]. These noises and changes deserve more efforts especially in spontaneous speech recognition because spontaneous speech is ill formed and usually includes redundant information such as disfluency, fillers, repetitions, repairs and word fragments [2].

The following sections describe recent trend in increasing performance of noisy speech recognition, with focus given to model compensation using Parallel Model Combination (PMC) method. This paper then discusses and compares several adaptation techniques. Finally, the concluding remarks are given in section 5.

2. Techniques for Dealing with Noise

Over the years, various techniques have been proposed by researchers to improve noisy speech recognition. These techniques fall into three main categories: noise resistance, speech enhancement and model compensation.

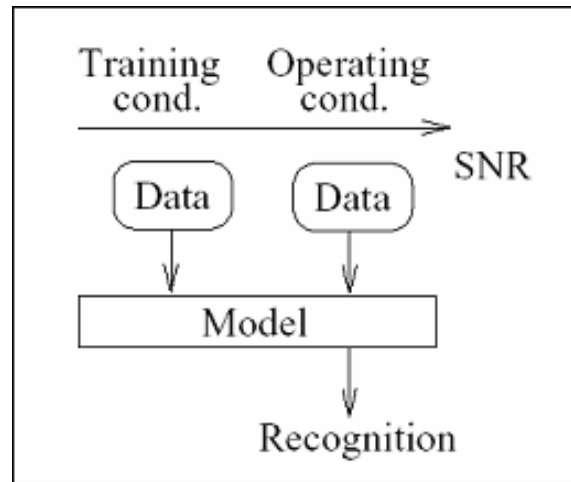


Figure 1. Noise resistance approach

Noise resistance (Figure 1) focuses on the effect of noise to the parameters of a model. Since the parameters of recognizer are sensitive to disturbances such as noise in the environment, by observing the effects of these disturbances, it is hoped that parameters that are noise resistant can be derived. In other words, the goal of this approach is to choose speech parameters so that the distortions due to noise are minimal. Recent researches of noise resistance techniques focus on developing computational models of the auditory system. Other techniques in this category include RASTA (Relative Spectral Processing), CMN (Cepstral Mean Normalization), linear discriminant analysis, etc.

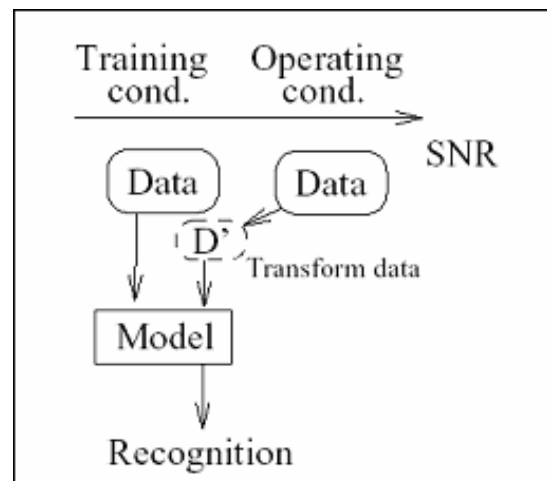


Figure 2. Speech enhancement approach

Speech enhancement (Figure 2) is usually used as a preprocessing step before recognition. The aim of speech enhancement is to restore clean speech signal or estimate the clean speech from the noisy speech. This involves the transformation of noisy speech into an approximation of clean data used in the training environment. This approach is originally introduced for speech quality improvement. In this case, performance can be assessed by means of intelligibility, quality measures, and distortion between

clean and recovered speech. When it is used in speech recognition, performance is mainly in terms of recognition accuracy. Depending on the application domain, it is important to select the relevant evaluation measure because some techniques may enhance the speech but reduce the recognition accuracy [6]. Spectral subtraction, statistical modeling, mapping transformation, etc. are techniques based on speech enhancement.

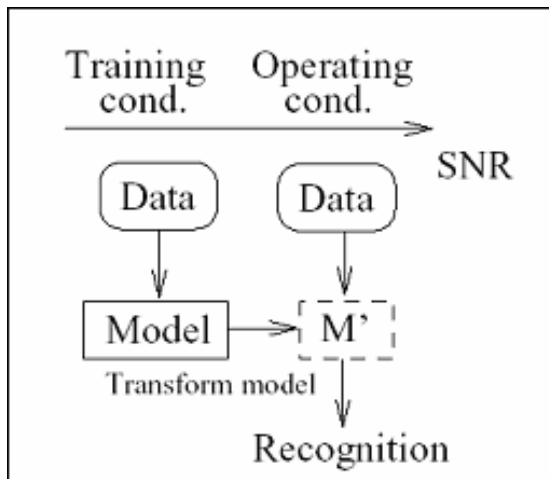


Figure 3. Model compensation approach

As mentioned previously, the degradation of performance is mainly caused by difference between the training and operating environment. The most straightforward approach to obtain good recognition models is to collect enough data from the operating environment, and use the data for training. Anyway, this is almost implausible and very time-consuming. As a result, the recent approaches for noisy speech recognition trend to utilize limited data from noisy environment to update the clean speech models [7]. This approach is called model compensation (Figure 3). Unlike other approaches, model compensation allows the presence of noise in the recognition process. Usually, Hidden Markov Models (HMMs) are used as the framework of this approach. Model Compensation entails the transformation of the parameters of a clean speech model to accommodate noisy speech. The transformation on the model parameters reduces the discrepancy between training and operational conditions, thus decreases the mismatch between the training and testing. Noisy parameters such as mean and variance are usually adapted for the compensation purpose [6]. Since model compensation modifies the acoustic models in the pattern matching stage instead of the parameterization stage, there is an advantage that no decisions or hypotheses about the speech are necessary and the observed data is unaltered [8]. Techniques for model compensation include HMM decomposition, Parallel Model Compensation (PMC), data contamination, Minimum Error Training, adaptation of HMM parameters, etc.

3. Parallel Model Combination

Among the model compensation techniques, PMC has been one of the most popular used techniques. The aim of PMC is to alter the parameters of a set of HMM Based acoustic models, so that they reflect speech spoken in a new acoustic environment. Figure 4 shows the basic process of original PMC method [8]. The first step of the process is to transform the inputs, i.e. the clean speech models and noise model, into the linear-spectral domain or log-spectral domain. Transformation is needed because most state-of-the-art speech recognition systems are based on cepstral parameters, but the effects of noise are best modeled in the linear spectral or log-spectral domain. In these domains, the clean speech model and noise model are combined using some mismatch function, which is the function that describes how noise affects the speech parameters. After the model combination, the corrupted speech model is transformed back into the cepstral domain for normal recognition process.

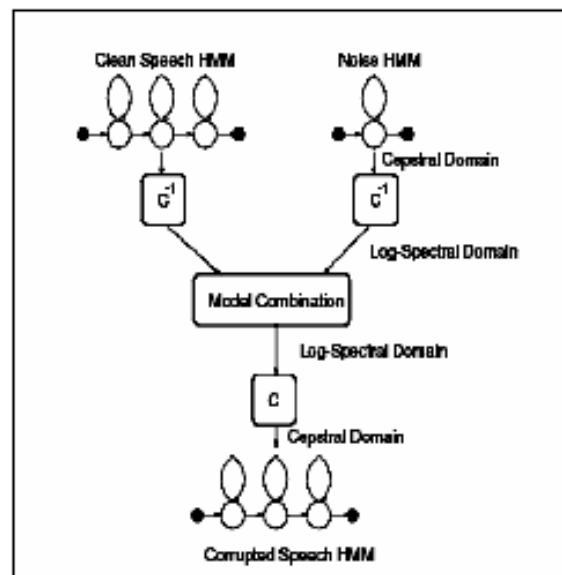


Figure 4. The basic process of original PMC method

In the MFCC domain, PMC has been shown to produce greater improvements in recognition rate than that of Varga's HMM decomposition technique [9]. Comparable results were also achieved for non-linear spectral subtraction, but MFCC model combination was predominant because the technique is applicable to a wider range of noise environments and it is not dependent on accurate end-pointing during recognition [10]. PMC was also shown to be robust against additive noise for SNRs as low as 0dB in the cases of stationary and non-stationary noises. In Vaseghi and Milner [11], good performance has been reported for impulsive noise such as machine gun noise.

The original PMC method is the log-normal PMC [12], which uses log-normal approximation [10] during model parameters estimation. Unfortunately the computation requirements are still relatively high, so a simplified version

of PMC which uses log-add approximation [13] is introduced. Anyway, this approach is limited, as it may not be used to compensate the variance parameters.

Another variant of the PMC method is the data-driven PMC (DPMC) [13]. In DPMC, the Gaussian integration to estimate the parameters of corrupted speech models is replaced by the generating of corrupted speech observations. That is, a speech observation and a noise observation for a particular pair of speech and noise states are generated, and then combined according to the appropriate mismatch function. Finally, a maximum likelihood estimate of these noisy samples is computed. DPMC is faster and it is also less computationally expensive compared to PMC with comparable performance. However, study showed that the accuracy of DPMC highly depends on the number of samples used to train the noise-corrupted speech model, thus increases the computation complexity while improving the performance [7]. It is also believed that some assumptions and approximations made in the PMC, primarily in the domain transformation and parameter combination processes, are not accurate enough in certain practical situations. Three new approaches, including the truncated Gaussian approach and the split mixture approach for domain transformation process and the estimated cross-term approach for parameter combination process, are proposed for improvement [7]. Experiments showed that the proposed approaches are able to provide significant improvements over the original PMC method, especially when the SNR condition is worse. It is believed that the proposed approaches can be further improved if the various parameters of the three approaches can be set differently for different models or different mixtures of the models.

Sarikaya and Hansen [14] claimed that although PMC is effective, the intense computational complexity limits its use in real-time use. Therefore, they presented a new framework for fast model combination, by using principal component analysis (PCA) to exploit a priori knowledge of the clean speech models and using efficient rectangular DCT (discrete cosine transform) and inverse DCT matrices for domain transformation. Keeping the same performance as original PMC, the new PCA-PMC method has increased the computational speed by a factor of 1.9. A direct adaptation scheme of the cepstral variance using a linear interpolation has also been introduced by Hwang, Yuo and Wang [15] to avoid the mapping from linear-spectral domain to the cepstral domain. The interpolation weight is effectively a similarity measure of the combined signal to the clean speech. Two methods to compute the weight is introduced by either using an energy ratio or a distance ratio. This method provides a comparable or superior recognition accuracy to the original PMC and a lot of computation resulted from the parameter mapping can be reduced. Therefore, the improvement of this method is twofold. Since a great deal of computational complexity of PMC lies behind the domain transformation step, Kim and Rose [16] stated the potential use of cepstrum decomposition technique to the PMC framework, so that the corrupted speech models can be obtained by adding the

means and variances of clean speech models and estimated noise models.

Another shortcoming of PMC is that, it is necessary to find the silence periods from utterances to sample the noise signal and estimate the parameters of the noise model. However, it may not be possible to obtain sufficient noise samples to estimate reliable noise parameters in the short periods of silence. Furthermore, the obtained noise samples may fail to represent the true statistics of the noise signal if the noise changes instantaneously. Therefore, Chung [17] suggested using an expectation-maximization (EM)-based adaptation approach for the mean of noise. This method has shown improved results compared with the conventional PMC. Instead of assuming statistical approximations as in the conventional PMC, the various statistical information necessary for the HMM parameter adaptation can be directly estimated by using Baum-Welch algorithm. Chung [18] proposed a data-driven adaptation method for the HMM parameter compensation. Generally, the method showed improved performance compared to the conventional PMC. Improvement was especially significant at low SNRs.

PMC has also been used along with Mel Frequency Discrete Wavelet Coefficient (MFDWC) features to take advantage of both noise compensation and speech features local in frequency domain [19]. It was shown that the MFDWCs are superior to the MFCCs for speech recognition in adverse environments when they are used in conjunction with the PMC method.

More recently, Novotny, Sovka and Uhlir [20] proposed an original method combining log-add PMC with a noise power spectral density estimation based on minimum statistics. The stability of speech vector selection under the influence of various background noises has made the suggested solution a beneficial method. The main advantages of this solution are no extra voice activity detector and a relatively low computational load.

From the review above, it is apparent that efforts that had been done on PMC attempted to accomplish two tasks; first, to improve the accuracy of noisy speech recognition, and second, to save the computation and the costs. In balancing these two aspects, adaptation technique is the approach that gives a twofold reward.

4. Adaptation Techniques

A large amount of data is often needed in order to obtain sufficient model accuracy during training of acoustic models. But the collection of data is time-consuming and expensive. As an alternative, adaptation techniques can be applied for adaptation of HMM parameters using a small amount of adaptation data. Most of them were originally introduced for speaker adaptation, i.e. to adapt the models to a particular speaker or a group of speaker. However, these techniques can also be used to perform environmental adaptation by using only a small amount of data from the new acoustic environment (i.e. the noisy environment in our context) and thus reducing the computation and the costs.

Besides, adaptation technique had been proved effective in improving the accuracy of recognition (e.g. [8], [21]). The following table shows the Word Error Rate (WER) of using PMC alone, and when applying PMC with an adaptation technique (Maximum Likelihood Linear Regression or MLLR).

Table 1. Performance when using PMC alone and when applying PMC with MLLR (From [21])

Method	Word Error Rate (WER)	Improvement
PMC	7.7	-
PMC + MLLR (mean adaptation)	6.7	12.99 %
PMC + MLLR (mean adaptation and variance adaptation)	6.6	14.29 %

Adaptation of HMM parameters can be done using several methods such as Maximum A Posteriori (MAP) [22] and Maximum Likelihood Linear Regression (MLLR) [23]. The benefits of these techniques can be evaluated from two aspects: first, how effective is the adaptation; second, the amount of adaptation data needed for effective adaptation. MAP (sometimes referred to as Bayesian adaptation) involves the use of priori knowledge about the model parameter distribution. Hence, if the priori knowledge can give information about what the parameters of the model are likely to be, it is possible to use limited adaptation data to obtain a satisfactory MAP estimate. One obvious drawback to MAP is that it requires more adaptation data to be effective when compare to MLLR. However, when large amounts of adaptation training data become available, MAP begins to perform better than MLLR. MLLR is a transformation-based technique and it works by computing a set of transformations to alter the means and variances in the initial model, so that the model is more likely to generate the adaptation data. MLLR usually updated only the means and variances because the main differences of speakers or environments are assumed to be characterized by these parameters.

More recently, an unsupervised incremental adaptation technique is proposed for rapid adaptation, which adapts the recognizers to a particular user or environment as the systems are being used. This technique is called Maximum Likelihood Spectral Transformation (MLST) [24]. MLST has some advantages over both the MLLR and the MAP. MLST has small number of parameters to be estimated, so it requires only a small amount of adaptation data. MLST is also a transformation-based technique and the transformation is done in the linear-spectral domain so that the adaptation process is computationally inexpensive. Besides, Kim and Yook [24] also proved that MLST produces a lower Phone Error Rate (PER) than MLLR:

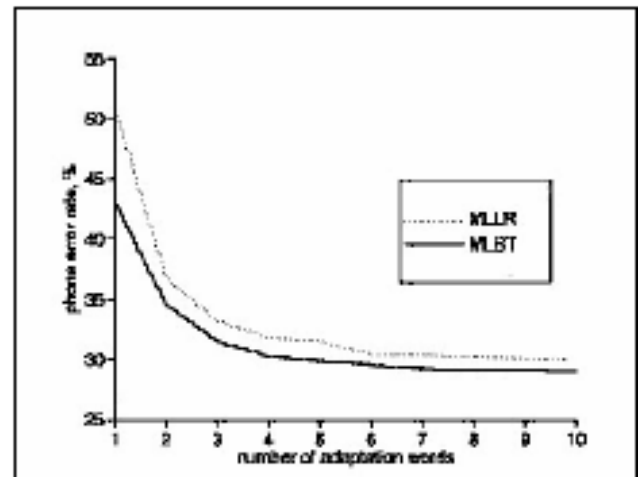


Figure 5. Average PER of MLLR and MLST (Adapted from Kim and Yook [24])

Hence, in this research, MLST will be used to adapt the parameters of the corrupted speech HMMs. Figure 6 shows the framework of the suggested method, which is denoted as Adaptive Parallel Model Combination (APMC) in this research.

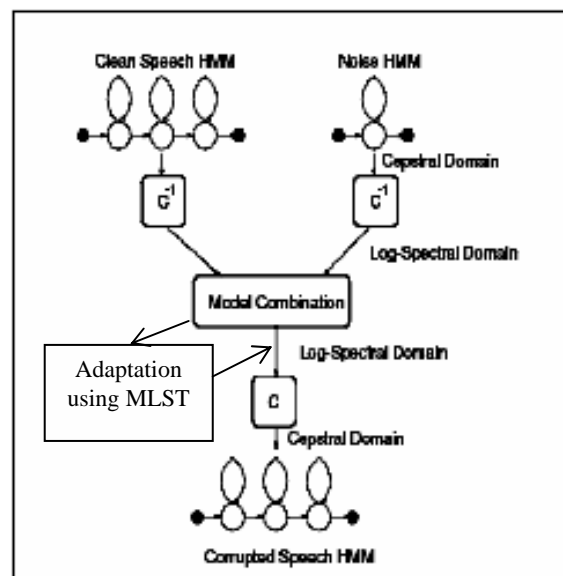


Figure 6. Framework of the suggested method

5. Conclusion

In this paper, various approaches to improve noisy speech recognition have been discussed. PMC has been one of the most successful techniques. However, there are several shortcomings on the conventional PMC, primarily in the domain transformation and parameter combination processes. Besides of the extra computational complexity, some assumptions and approximations made in these processes are not accurate in certain situations, thus restricting better performance of PMC. A lot of methods have been proposed to deal with these problems; e.g. using PCA, truncated Gaussian approach, estimated cross-term

approach, etc. In this paper, a method denoted as APMC, which combines PMC and MLST has been proposed. MLST will be used to adapt the parameters of the corrupted speech HMMs produced by the conventional PMC to further increase the recognition accuracy of PMC method.

Reference

- [1] Hunter, A. J. S. (2002). Mobile GIS as if Field Users Mattered: Small is Ubiquitous But Can Speech Be Recognized. Master Thesis, Department of Geometrics Engineering, University of Calgary.
- [2] Furui, S. (2005). Toward Robust Speech Recognition and Understanding. *Journal of VLSI Signal Processing*. 41: 245-254.
- [3] Oviatt, S. (2000). Taming Recognition Errors With a Multimodal Interface. *Communications of the ACM*. 43(9): 45-51.
- [4] Rose, R. (2004). *Environmental Robustness in Automatic Speech Recognition*. McGill University.
- [5] Markowitz, J.A. (1996). *Using Speech Recognition*. 1st ed. United States of America: PTR Prentice Hall. 35-47, 50-52, 146-160.
- [6] Junqua, J-C. and Haton, J-P. (1996). *Robustness in Automatic Speech Recognition*. 1st ed. United States of America: Kluwer Academic Publishers. 736-113, 155-180, 273-312.
- [7] Hung, J-W., Shen, J-L. and Lee, L-S. (2001). New Approaches For Domain Transformation and Parameter Combination For Improved Accuracy in Parallel Model Combination (PMC) Techniques. *IEEE Transactions on Speech and Audio Processing*. 9(8): 842-855.
- [8] Gales, M. J. F. (1995). Model-based Techniques For Noise Robust Speech Recognition. Ph.D Dissertation, University of Cambridge.
- [9] Varga, A. and Moore, R. (1990). Hidden Markov Model Decomposition of Speech and Noise. In *ICASSP*. 845-848.
- [10] Gales, M. J. F. and Young, S. (1992). An Improved Approach to the Hidden Markov Model Decomposition of Speech and Noise. *IEEE*. 233-236.
- [11] Vaseghi, S. and Milner, B. (1995). Speech Recognition in Impulsive Noise. In *ICASSP*. 437-440.
- [12] Gales, M. J. F. and Young, S. (1993). Cepstral Parameter Compensation For HMM Recognition. *Speech Communication*. 12(3): 231-239.
- [13] Gales, M. J. F. and Young, S. (1995). A Fast and Flexible Implementation of Parallel Model Combination. *ICASSP*. 131-136.
- [14] Sarikaya, R. and Hansen, J. H. L. (2000). PCA-PMC: A Novel Use of a Prior Knowledge For Fast Parallel Model Combination. *IEEE*. 1113-1116.
- [15] Hwang, T-H., Yuo, K-H. and Wang, H-C. (2001). Linear Interpolation of Cepstral Variance For Noisy Speech Recognition. *Eurospeech*.
- [16] Kim, H.K., and Rose, R. (2003). Cepstrum-Domain Acoustic Feature Compensation Based on Decomposition of Speech and Noise For ASR in Noisy Environments. *IEEE Transactions on Speech and Audio Processing*. 11(5). 435-446.
- [17] Chung, Y-J. (2002). Adaptation Method Using Expectation-Maximization For Noisy Speech Recognition. *Electronics Letters*. 38 (13). 666-667.
- [18] Chung, Y-J. (2005). A Data-Driven Model Parameter Compensation Method For Noise-Robust Speech Recognition. *IEICE Trans. Inf. & Syst*. E88-D(3): 432-434.
- [19] Tufekci, Z., Gowdy, J.N., Gurbuz, S. and Patterson, E. (2001). Applying Parallel Model Compensation With Mel-Frequency Discrete Wavelet Coefficients For Noise Robust Speech Recognition. *Eurospeech*.
- [20] Novotny, J., Sovka, P. and Uhler, J. (2004). Study and Application of Silence Model Adaptation For Use in Telephone Speech Recognition System. *Radioengineering*. 13(3). 1-6.
- [21] Gales, M. J. F. and Woodland, P. C. (1996). *Variance Compensation Within the MLLR Framework*. University of Cambridge.
- [22] Gauvain, J.L. and Lee, C.H. (1994). Maximum A Posteriori Estimation For Multivariate Gaussian Mixture Observations of Markov Chains. *IEEE Trans. Speech Audio Process*. 291-298.
- [23] Leggetter, C.J. and Woodland, P.C. (1995). Maximum Likelihood Linear Regression For Speaker Adaptation of Continuous Density Hidden Markov Models. *Comput. Speech Lang*. 171-185.
- [24] Kim, D. and Yook, D. (2004). Fast Channel Adaptation For Continuous Density HMMs Using Maximum Likelihood Spectral Transform. *Electronics Letters*. 40(10).

A Perspective of Manufacturing Features from Islamic Views

Setyawan Widyarto¹, Muhammad Saffie bin Md Yasan²,
Mohd Fahmi Mohamad Amran² and Habibollah Haron²,

¹ Politeknik Negeri Lampung,
Jalan Soekarno Hatta, Rajabasa, Bandar Lampung 35145, Indonesia
Tel. 62-721-703995, Fax. 62-721-787309, E-mail: swidyarto@lycos.co.uk

² Department of Modeling and Industrial Computing
Faculty of Computer Science and Information System,
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor
Tel.: +60 (07)-5532070, Fax: +60 (07)-5565044, E-mail: saffie@kuitho.edu.my, fahmi.mohd@gmail.com

Abstract

The study aims to make a balance between existing manufacturing system and Islamic management views. Besides some review of existing works done, some day to day basis experiences also decorate the methods of presentation. This paper discusses two existing main issues in shop floor that lead conflict in the past and potentially still going on. The first is an issue of working time in manufacturing and the second one is an issue of clothing in work place. The discussion will focus on 24 hour-working time and 'inhuman' working hours. Working hours would be included 'inhuman' when the working time hinders workers from doing other needs that may not be replaced. The review of non-stop manufacturing, its advantages and disadvantages will be exposed. Certainly, round-the-clock operation still makes economic sense. However, nowadays' customers are demanding increasingly high-performance, high-quality products and shorter lead times. Studies of round-the-clock operations show that there are significant hidden costs associated with workforce changes that occur every eight or twelve hours. Available evidences indicate that improving working time arrangements could potentially promise significant increase of profit. With such arrangements, workers can improve their mental and physical health and thus be more rested and alert during working hours, thereby improving labour productivity (such as reducing errors, avoiding absenteeism) and dropping health care costs. In this discussion, clothing issues at work place are limited to special requirement by management in such a way the workers have to disobey their faith. The fact that more women are entering global labour force in manufacturing and services sectors cannot be stopped. The issue of [woman] workers wearing special dress according their faith might emerge as hindrance. The paper perceives this issue.

Keywords:

Inhuman working hours, clothing at work place, labour productivity

1. Introduction

This paper's motivation is shift work pros and cons discussion, and alternative improvement of from Islamic views are proposed. The advanced technology has made work year-round becomes easier to find, since labour force was not so tied to the season, sun lighting, that artificial lighting made work possible for the greater part of the day. However, there has been work force migration that advents new socio-industrial problem. Peasants, often manipulated into positions of debt and disadvantage by individuals, have moved from the farms to the factories to work at labour for long periods of time. Some countries would have been able to overcome this problem by agricultural mechanization and machineries application. Some countries have to hire

labours from other countries to adequately supply their work force lacks. For examples, Malaysia and Singapore have to bring labours from their neighbour countries. Unfortunately, some individuals take opportunities from this situation and illegal immigrant sometimes becomes domestic and regional political issues.

Technological advances make working hours possible to extract upwards of seventy hours per week of working time from labours or twelve hours a day while factories are running 24-hour non-stop operation. Before collective bargaining and worker protection laws, there was a financial incentive for a company to maximize the return on expensive machinery in spite of the suffering of workers. Records indicate that work schedules as arduous as twelve to sixteen hours per day, six to seven days per week, were demanded of wage earners. However, the working hours declined because of rising wages brought by economic growth, with a supporting role from trade unions and collective bargaining, and progressive legislation. For example, France adopted a

* Corresponding Author. E-mail: swidyarto@lycos.co.uk,
Tel. 62-721-703995, Fax. 62-721-7873091

35-hour workweek in 2000. Working hours in industrializing economies like South Korea, though still much higher than the leading industrial countries, are also declining steadily. It is clear that labour force is merely a variable of manufacturing in respects of working hours and is very vulnerable against profit orientations.

2. Shift Work - Pros and Cons

Working time refers to the period of time that an individual spends at paid occupational labour. Unpaid labours such as housework are not considered part of the working week. Many countries regulate working time by law, such as stipulating minimum daily rest periods, annual holidays and a maximum number of working hours per week.

According to the ILO [1], working in shifts is “a method of organization of working time in which workers succeed one another at the workplace so that the establishment can operate longer than the hours of work of individual workers” at different daily and night hours. Within the broad category of shift work, we can make a distinction between a fixed shift system and a rotating one. Under a fixed shift system, working time can be organized in two or three shifts: the early, late and/ or night shifts. This means that one group of workers might work during the morning and early afternoon; another group will work during the late afternoon and evening; and the third group (in a three-shift system) will work during the night. Under a rotating shift system, workers might be assigned to work shifts that vary regularly over time; these are called “rotating shifts” because they rotate around the clock (e.g. from a shift in the morning, to one in the afternoon, to one at night). If the firm’s equipment is running non-stop the whole week, we can speak of “continuous” shift work with 24 hours of work, seven days a week (i.e. 168 hours of firm operating time). There is a virtually unlimited number of potential shift work patterns.

Shift work is typically introduced in response to business needs. Sometimes working overtime is not sufficient to meet demand. Alternatively, overtime may become so regular that it becomes inefficient. For example, an industry may experience an increase in demand but does not view the increase as long term. Therefore, it is more expedient to introduce a shift system temporarily. If over time need becomes unending, the company may decide to make the shift computer manufacture industry or the pharmaceutical industry, it can be more cost effective to keep the process running continuously than to stop and start in which case they may introduce a permanent shift system.

There are valid business reasons for introducing shift work but a number of advantages and disadvantages for the company may accrue.

Advantages for the company include more cost effective than overtime, increased production capability, reduced overheads e.g. off peak rates on services - electricity, telephones, and meeting additional demand without permanently increasing headcount (in temporary situations).

Disadvantages for the company include potential increase in absenteeism levels, turnover, de-motivation levels and sickness or ill health due to changes in employee metabolism from working night shift for a long period.

2.1. Hidden Cost of Inhuman Working Hours and Shift Works

When 24-hour operations first became popular, the main focus was to never shut down the process but today’s customers are demanding increasingly high-performance, high-quality products and shorter lead times. This means that running operations longer is no longer sufficient but they must also be run better. Studies [2] of round-the-clock operations throughout the United States show that there are significant hidden costs associated with workforce changes that occur every eight or twelve hours. These costs are quantifiable and often run in the range of \$4,000 to \$8,000 per employee per year. Therefore, a moderate-size manufacturing plant with 300 employees is probably absorbing between \$1.2 million and \$2.4 million a year in costs associated with poor process continuity, poor teamwork, ineffective communications and inconsistency in operating procedures. **Consider these situations:**

At one consumer products plant, process improvement was slow and temporary because of poor inter-shift communications. Achieving the site’s aggressive performance goals was in serious jeopardy and hinged on good coordination and collaboration across shifts. Profit Improvement Potential: \$2.3 million a year.

Despite the presence of standard operating procedures, such as ISO certification, each shift, in practice, that is operated independently of the others is more important to the competitive advantage and long-term success of the business. Not only was product quality inconsistent across shifts, but also valuable time was lost at each shift change for machine set-up and readjustment. Profit Improvement Potential: \$1.2 million a year.

At one chemical manufacturer, poor cross shift teamwork and coordination was causing long cycle times and reducing plant operating capacity. Millions of sales was being lost because of this inability to meet demand and streamline throughput across all shifts. Profit Improvement Potential: \$1.6 million a year.

These plants are in different industries and have different organizations and priorities, but they share something in common. In each, managers struggle with the issues of 24-hour management on a daily basis. And in each, there is significant opportunity to improve productivity and profitability. Some of the key costs and risks preventing U.S. companies from achieving the full potential benefits of extended hours operations, according to the study [3], are:

Lost productivity - Measured as output per employee hour, productivity is 5 percent lower between the hours of midnight and 7:00 a.m. than during the day. By using a model that estimates the value of goods or services that an extended hours employee must add to an operation (based on salary, benefits and overhead), the annual cost of lost night time productivity is \$79.4 billion, or \$3,309 per extended hour employee.

Absenteeism - Average absenteeism among companies with extended hour operations is more than twice the national average, at 4.9 percent versus 2.1 percent. The average incremental cost of overtime and temporary workers to cover for absenteeism is \$2,102 per extended hours employee, or \$50.4 billion annually, versus an average of

\$970 per non-extended hour employee.

Turnover - Average turnover is nearly three times higher among extended hour workers, with a turnover rate of 9 percent in 2002 compared with a national average of 3.4 percent. The average cost of recruiting and training extended hours employees is \$25,000, including overtime, lost productivity, administration, training, recruiting and separation costs, bringing the incremental cost to \$39.1 billion, or \$1,631 per extended hours employee.

Health care costs - Employees staffing extended hours operations suffer from significantly higher rates of obesity, gastrointestinal disorders, cancer, sleep disorders and fatigue-related car accidents. The study estimates that health problems exacerbated by continuous operations cost U.S. employers an additional \$28.1 billion annually, or \$1,181 per extended hours employee, in higher health insurance costs.

2.2. Shift Work and Its Health Risk

According to a range of studies summarized in the recent ILO report, *Working time: Its impact on safety and health*,[4] the most crucial factors affecting employees' health are long working hours and certain shift-work patterns, particularly those involving the irregular distribution of working hours and work at night. The report states that regular working in excess of 48 hours per week constitutes an important occupational stressor, which significantly increases the risk of mental health problems. Regularly working more than 60 hours per week clearly increases the risk of cardiovascular disease. There is strong evidence that shift work provokes cardiovascular and gastro-intestinal disorders.

Sleep disorders are also widely reported among shift workers, and there is some evidence of reproductive disorders caused by shift work. Night work disturbs the circadian rhythm in the human body. The risk of health disorders is further increased by shift work

- Increased heart disease
- Gastric ulcers and gastro intestinal problems
- Social problems and minor psychiatric disorders
- Sleep disorders and increased fatigue
- Increased error rates and accident rates.

Figure 1 shows the effects of a 65 hour week combined with 20 hours work travel on an individual's life-work time balance (from Table 1).

Table 1. Healthy vs. unsafe life-work time balances

Activities	Healthy	Unhealthy
Sleep	49	35
Household	35	28
Work	35	65
Travel to work	5	20
Family & friends	20	10
Leisure & fitness	20	10
Citizen	4	0
Total hours/week	168	168

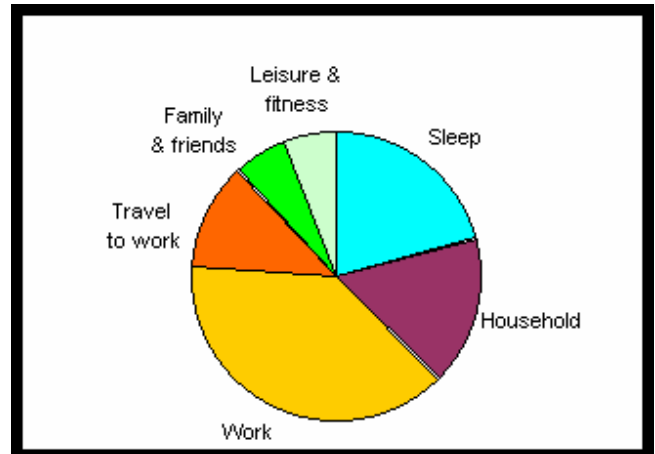


Figure 1. Unhealthy life-work

In this example work dominates 50% of the individual's life invading sleep, family and fitness time. Due to fatigue leisure time is more likely to be TV than exercise. There is little time for family and friends and none for citizen contributions (the voluntary work that communities rely on).

This hazardous life-work imbalance may have serious effects on physical health, mental health and relationships, day to day and cumulatively. Sooner or later these personal life effects will also have repercussions back across the life-work boundary. Life-work conflicts degrade motivation and may cause severe stress jeopardising work performance and costing absence for personal sickness or family crises

3. Manufacturing in Islamic View

Human has been already clarified as one of resource in production and many studies have been conducted to create more productivity from the human resources. Generally, income is a main motivation to control human [labour] productivity. A company may play around with bonus and incentives to motivate workers. However, performance-related pay and bonus schemes can significantly improve worker satisfaction levels as well as productivity but it works in beginning time of the implementation. Studies show that monetary incentives do not motivate as well as non-monetary incentives and monetary rewards have less lasting value. It has been found that it would take double the money to make a reward as worthwhile as any other incentive reward, travelling reward for instance. There is an indication that working time must be accompanied by off-duty time.

3.1. Potential productivity effects of working time change

The available evidence indicates that improving working time arrangements can have productivity-enhancing effects, which are similar to the effects of higher wages. With such improvements, workers can improve their mental and physical health and thus be more rested and alert during working hours, thereby improving labour productivity (such as reducing errors). Further productivity gains can be expected by improving staff recruitment, retention, motivation and commitment. Three mechanisms (among others), through which better

working time arrangements can translate into higher productivity, are particularly important [5]:

Physiological effects. The fundamental connection between working time and productivity exists at the physiological level. For example, it is known that long hours are accommodated by an adjustment of pace or work intensity, such as a slowing of movements and inserting more pauses between movements. Thus, the extension of working hours may not yield the expected increases in total output, and may possibly lead to a reduction of total output in the long run. *In Islamic manufacturing, pauses can be inserted according prayer time.* Conversely, the reduction of working hours can be at least partly offset by increases in the pace of work without causing physically harmful consequences.

Motivational effects. Better working time arrangements can have motivational effects by obtaining the psychological willingness of workers to use their energy in more efficient ways. Such motivational gains, however, might be lost if management does not recognize them and takes no steps to change working time practices and methods to make use of and secure these gains. It is also known that these effects can only be expected if there is a positive cooperation between managers and workers in an enterprise, which provides a facilitating environment for a positive reciprocal relationship between better working time arrangements and higher labour productivity. *Islam teaches human to start work once Shubuh prayer finish. Before the sun rises, workers have arrived their work place. Operators may start his machine or a factory may initiate running exactly when the sun rises.*

Improved organization of work. Changes in working time can induce a series of productivity-enhancing changes in work organization. They may invite management to identify and reduce “non-productive” time that arises due to inefficiencies in scheduling and methods of supervision. *Provided, there are five times of prayer and management regularly let workers pause and hold very short briefing and evaluation of the daily productivity, any errors and target behind schedule can be detected early and corrected at the first opportunity.*

3.2. Work & Family

The interacting trends in workplaces and in families can lead to considerable conflicts and stress for workers and their families, as well as negative consequences for the workplace and society as a whole.

For families: The declining availability of family support for care responsibilities, combined with increasing pressures in the work environment (including insufficient and insecure incomes), is putting considerable strain on families. For childcare, families may be reduced to less than satisfactory solutions, such as leaving children alone at home, taking an older child out of school or bringing the child to work where the environment may be unhealthy or dangerous. For example, in India it is estimated that 7 or 8 million children accompany their parents (mostly mothers) on building sites.

For workers: Workers in many countries (both men and women) are feeling increasingly stressed by the competing demands of work and family for their time and their energy. The pressures and stress are particularly high for women

workers, who typically continue to assume responsibility for domestic tasks and care of dependents. Dealing with family emergencies — such as a sick child or parent — can lead to absenteeism, for which pay may be lost and, in some extreme cases, the result may be dismissal.

For work: The problems which work family conflicts pose for workers inevitably impact on the workplace. Problems such as turnover and difficulties in recruiting appropriate personnel may be related to the incompatibility of working conditions with family responsibilities. Absenteeism and tardiness may be the result of work-family conflicts which could be avoided. The stress on staff with family responsibilities may be affecting their concentration and productivity. The costs of these problems have often not been assessed in relation to the costs of adopting more family-friendly arrangements. In fact, research suggests that many enterprises are just unaware of the business benefits that can be acquired by adopting a more family-friendly stance.

For societies: Work-family conflicts can have general effects on society as a whole:

- Perpetuating poverty and the inequities within society, since family responsibilities can constrain possibilities for income generation, particularly for low-income families who cannot afford paid solutions such as domestic help or day care;
- Creating difficulties for the provision of appropriate care for dependent children and the elderly;
- Preventing the labour force participation of women and the use of their education and skills;
- Contributing to a decline in fertility by discouraging parenthood (particularly motherhood); and
- Preventing equality of opportunity and treatment for men and women in employment, since the constraints and disadvantages that family responsibilities bring in the labour market fall mainly on women.

In Convention No. 156, “family responsibilities” refers specifically to responsibilities in relation to “dependent children” and “other members of the immediate family who clearly need their care”. It is up to each country to define which persons would be covered by these terms, given that the notions of “family” and “family obligations” can take many forms in different societies and contexts.

Indeed, a labour besides as a employee he/she is also a family member in which he/she has duty to fulfil family members' right. Q.S. Al Israa'(17). 26: And give kinsmen their right and to the needy and the traveller and spend not extravagantly.

3.3. Quality of Life

Working must be balance with other activities such as spending time with family, living in community, leisure and fitness activity. Islam declares human are is not perfect and also needs relationship between human and the God, Allah, social lives amongst them and interactions with environment (including animal and plant).

Q.S. Adz Dzaariyaat.(51). 56: And I (Allâh) created not the jinns and humans except they should worship Me

(Alone).

Q.S Al Hujuraat.(49).13: 'O people! We created you from one man and one woman and made you branches and tribes that you may recognize one another. Undoubtedly, the most respected among you in the sight of Allah is he who is more pious, verily, Allah is knowing, Aware.

Q.S. Thaahaa (20).53: He Who made the earth for you as a bed and laid for you walking paths in it and sent down rain from the sky. And thereby We brought forth-varied pairs of vegetation.

Human is so weak that Islam shows and teaches human how to organize their lives. "God wants to make easy for you; and mankind was created weak" (Q.S.An-Nisa` : 28)

Night and day creation for human life is foundation to spend time. "Initiator of morning and Maker of the night to reside in; and the sun and the moon for counting. Such is the measure of the Noble, the Knowledgeable" (Al-An'am: 96). "He is the One who made the night for you to reside in, and the day to see. In that are signs for a people who listen" . "And He is the One who made for you the night as a cover, and sleep for resting, and He made the day to move about in" (Al-Furqan: 47)

Truly, Allah made night for human sleep and rest after along morning people work. This approve by researcher6 about important sleep for quality of life. In their study, To maintain a healthy life-work balance we need about half our time for sleep and essential household tasks (see Figure 2 and Table 1). Total the sleep and household are 84 hours. That means human needs no less than 12 hours per day for activities out of working.



Figure 2. Healthy life-work balance

The biggest issue on the life-work boundary is time. Time management is important to employers. It is equally important to employees. Human needs start with sleep (at least 7 hours a day) and essential household tasks - food, hygiene, health care, shopping, maintenance and personal administration. In the past household tasks could be shared with an unwaged partner based at home. Now, single people and working couples need an average of 35 hours a week each for household tasks [6].

The model working week is about 35-37 hours (9-5 with lunch) over 5 days, plus about an hour a day travelling time. In practice 25% of UK employees work over 48 hours. Retail sector staff are often expected to work 8-9 hours x 6 days a

week. Transport operators and 24 hour operations (manufacturing, hospitals etc) want maximum use of facilities and expect 10-12 or even 24 hour shifts.

Table 1 contrasts healthy and unsafe life-work time balances. To maintain a healthy life-work balance we need about half our time for sleep and essential household tasks. This leaves the other half (84 hours a week) for work including travel and quality time for family, friends, leisure and fitness.

The EU Working Time Directive prescribes a maximum 48-hour working week. If additional work travel time is involved this makes serious intrusions into personal time - well over a healthy limit. But in many UK organisations it is nearer the norm. Many employers have persuaded staff to sign contracts to opt out of the EU Directive. Manual workers often rely on extra hours to earn overtime. Many white collar occupations e.g. managers and teachers and the self-employed also have to work additional time at home and weekends (2-4+ hours a day) giving a 60-70+ hour week.

Travelling time is taking an increasing share of UK working time. The high cost of house moves and short term employment contracts have led to increased commuting to avoid relocation costs, typically 2+ hours a day. In urban areas this may be further increased by congestion to 3+ hours a day and for travelling occupations to 5 or more hours a day. See Table 2.

Table 2: Life-work boundary issues: work culture and personal wellbeing Healthy vs. unsafe life-work time balances Formatting Rules

Life zone	Safe	Negotiable	Hazardous	Dangerous	Work Culture
Time	own & family	contracted hours	unpaid overtime	unlimited demands	Work & travel time
Money	sufficient for needs	opportunity incentives	low pay & insecurity	exploitation retaliation	Pay and benefits
Physical health	healthy & fit	energy & effort	fatigue & strain	accident or injury	Workload
Mental health	calm, fulfilled	stimulated, motivated	anxiety, stress	panic, anger or violence	Pressure, demands
Relationships	support & cooperation	respect, direction	over-control	harassment, abuse	Power, control
Identity (gender, race)	valued, shared	dignity, respected	devalued, prejudiced	scapegoating, excluded	Values & customs

3.4. Dressing in Workplace based on Islamic Views

Due to the changing ratio between men and women, women labour forces are not avoidable. Women are increasing entering the global labour force either in manufacturing or services sectors. Lately, the issue of women workers wearing hijab in the workplace got many attentions from various parties. Mostly Non-Islamic West Management has not supported the women who are wearing hijab when doing their works.

In one case at Belgium, Naimi Amzil who is Moroccan origin was forced to quit her job in the Remyery seafood factory on March 3, 2005 after receiving death threats from an extremist group for no reasons other than being a veiled Muslim [7]. It is pity because the one European country is reacting offensively to the Muslim people. Their got no respect and esteem to human religion.

Islamic dress is one of many rights granted to Islamic women. Islamic clothing is worn in obedience in the frame the relationship between human and the God. In many cases, Muslim women wish both to work and to maintain their religious convictions. It should be possible to fulfil both goals.

Muslim women report that the issue of attire comes up most often in the initial interview for a job. Some interviewers will ask if the prospective employee plans to wear the scarf to work. Others may inappropriately inquire about religious practices or beliefs.

Sometimes the prospective employee, feeling pressure to earn a living, will take off the scarf for the interview and then put it on when hired for the job. Modest dress should not be equated with incompetence. Other issues include unwanted touching or pulling on scarves by other employees, verbal harassment or subtle ostracism and denial of promotion.

In manufacturing scope, the advantages for Muslim women wearing hijab in the workplace is it can prevent Muslim women from any kind of harassment in workplace because they got respect from other workers for her attitude towards her religion. Besides, logically the probability Muslim women who are wearing hijab get head injured or hair problem are less because they cover their head using scarf.

Definitely, Islamic view of dressing in workplace is not limited to hijab. Workers wearing trousers that are not lower than the ankles are less risk when step on any staircase.

Exploiting women as front officers in any servicing enterprises with any non-Islamic values would mislead customers and business core. Instead of exposing 'body beauty', it is recommended the servicing enterprises to focus on main service quality measurement. For example, officers' communication skills, space comfort, varying menu, hygienic awareness, short waiting time, appropriate quick response to any complaints, and other customers' demand matching.

4. Conclusion

Several types of work may need shift working time because the jobs are very critical and related to security and human life such as hospital, communication, security and army. However, some types of work that operates work shift due to production target should be aware of hidden cost related. Alternatively, one long day time shift that practices several pauses or several short shift during day time only would be possible to apply.

References

- [1] International Labour Organization, 15 June 2006. Citing Internet Source, <http://www.ilo.org/public/english/>
- [2] Don Perkala, Granite Bay Consulting, 15 June 2006. Improving productivity & profitability, Citing Internet Source, <http://www.granite-bay.com/>
- [3] Sandy Smith (2003) – “Study: Shiftwork Operations Losing Major Profit Gains” Citing Internet on 15 June

2006. <http://www.occupationalhazards.com/articles/7389>

- [4] Anne Spurgeon (2003), International Labour Office “Working time: Its impact on safety and health”. Citing Internet on 15 June 2006, <http://www.ilo.org/public/english/protection/condtrav/pdf/wtwo-as-03.pdf>
- [5] M. White: Working hours: Assessing the potential for reduction, (Geneva, ILO, 1987).
- [6] International Labour Organization - “Working Time And Productivity” Citing Internet on 1 June 2006, <http://www.ilo.org/public/english/protection/condtrav/pdf/infosheets/wt-18.pdf>
- [7] Nasreddine Djebbi, Citing Internet on 11 June 2006, http://www.islamicawakening.com/viewnews.php?news_ID=4394&

Bibliography

- 1) Holy Quran, English Translation by Dr. Muhammad Taqi-ud-Din Al-Hilali, Ph.D., Dr. Muhammad Muhsin Khan
- 2) ACDM Newsletter (2000), “The Eos Life~Work Resource Centre: Briefing #2” - <http://www.eoslifework.co.uk/boundaries.htm>
- 3) Henrik Boggild (2003) – “Shift Work & Heart Disease – Epidemiological and Risk Factor”
- 4) Atanu Kumar Pati, Arti Chandrawanshi, and Alain Reinberg, Current Science Vol 81, No 1 (2003) – “Shift Work : Consequence and Management”.
- 5) J M Harrington, BMJ Journals (2001) – “Health Effects Of Shift Work And Extended Hours Of Work” - <http://oem.bmjournals.com/cgi/content/full/58/1/68>
- 6) Joseph Mercola (2001), “Shift Work Dangerous to Your Health” - www.mercola.com/2001/sep/29/shift_wor'k.htm
- 7) Mathen D. (University of Sydney, 1993) – “Shiftwork And Health. A Literature Review And Practical Approach To Shiftwork Roster Redesign.”
- 8) J. Heymann, International Labour Organization (ILO) , 2004 – “How are workers with family responsibilities faring in the workplace?”, <http://www.eldis.org/static/DOC16359.htm>
- 9) H. Gospel - “Quality of working life: A review on changes in work organization, conditions of employment and work-life arrangements (2003), Conditions of Work and Employment Series No. 1”
- 10) M. Abe, C. Hamamoto and S. Tanaka - “Reconciling work and family: Issues and policies in Japan (2003), Conditions of Work and Employment Series No. 5”.

A Physical Force Based Microscopic Crowd Model Simulation

Setyawan Widyarto¹, Aboamama Atahar, Muhammad Shafie Abdul Latif²

¹Politeknik Negeri Lampung,
Jalan Soekarno Hatta, Rajabasa, Bandar Lampung 35145, Indonesia
Tel. 62-721-703995, Fax. 62-721-787309, E-mail: swidyarto@lycos.co.uk

² Faculty of Computer Science and Information System,
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Email
Tel: +60-7-5535914, Fax: +60-7-5536688, E-mail: swidyarto@lycos.co.uk, khrfoto@yahoo.com, shafie@fksm.utm.my

Abstract

This paper presents a simulation of a microscopic crowd model. The model represents detailed interaction of pilgrims to control their movement in any pilgrim arena. The microscopic pilgrim flow characteristics need to be explored in controlling their movement. A physical based microscopic pilgrim simulation model has been developed. The research is a computer simulation model of pilgrim movement where every pilgrim member in the model is treated as an individual agent. The microscopic pilgrim simulation model has been developed to determine the microscopic characteristics of pilgrim flow.

The developed microscopic pilgrim simulation model is a physical force based model similar to the social force model with forward and repulsion forces as the main force driver. However, the detail of the model is somewhat customized with inputs of the model. The repulsive force is added to guarantee collision avoidance. The developed model also uses the physical based variables that can be measured. The collision avoidance algorithm is influenced by existing algorithm flock which uses the steering behaviour to animate birds. Thus, the developed microscopic pilgrim simulation model is made based on the existing models to improve the deficiency of the existing models and keeps their main advantages.

It was revealed that the microscopic crowd model studies have been successfully applied to explore to the behaviour of microscopic pilgrims flow by showing influenced three forces. Three forces are responsible to move ahead of individual pilgrim, move away for adjusting distance between two pilgrims, and prevent a collision when there are many pilgrims in the arena.

Keywords:

Microscopic crowd model, pilgrims, flow characteristics, physical force, collision avoidance

1. Introduction

Microscopic crowd model represents detailed interaction of pilgrims to control their movement in any pilgrim arena. The microscopic pilgrim flow characteristics need to be explored in controlling their movement. A physical based microscopic pilgrim simulation model has been developed. The features of the flow performance or microscopic pilgrim characteristics are explained.

The Microscopic Model and Simulation of Pilgrim is a computer simulation model of pilgrim movement where every pilgrim member in the model is treated as an individual agent. The microscopic pilgrim simulation model has been developed to determine the microscopic characteristics of pilgrim flow.

The developed microscopic pilgrim simulation model is a physical force based model similar to the social force model with forward and repulsion forces as the main force driver

[1]. However, the detail of the model is somewhat customized with inputs of the model. The repulsive force is added to guarantee collision avoidance. The developed model also uses the physical based variables that can be measured. The collision avoidance algorithm is influenced by Reynolds [2],[3] which uses the steering behaviour to animate birds. Thus, the developed microscopic pilgrim simulation model is made based on the existing models to improve the deficiency of the existing models and keeps their main advantages.

Pilgrims in the microscopic simulation model are modelled as Non-Player Characters (NPCs). NPCs are the autonomous characters that are free from the user's control but initial conditions are keyed by the users. NPCs are seen from above of the facilities (top view). A pilgrim is modelled as a circle with a certain radius. Each pilgrim's initial conditions includes initial location, initial time, and initial velocity, and predetermined target location (opposite to the initial location). These inputs can be besides determined by the user as a design experiment and be specified randomly.

Non-Player Characters (NPCs) are the autonomous characters that are free from the user's control. NPCs will

* Corresponding Author. E-mail: : swidyarto@lycos.co.uk,
Tel: +60-7-5532396, Fax: +60-7-5536688

interpret an action and this interpretation process is important for this view in behavioural animation. This will lead to further autonomous actions in the virtual environment as well as intelligent responses to the action being carried out. Thus motion or path planning becomes much more complicated when an animation for large crowds must be made. The development in motion planning and in global techniques for improving the approach has been discussed [4] but it concentrated on the probabilistic roadmap (PRM). Whereas the improvement for path planning techniques used for large crowds is very few.

2. Modelling Pilgrimage Movements

Briefly, the method used is depicted in Figure 1. However, this paper is going-on result and the fully steps have not conducted yet.

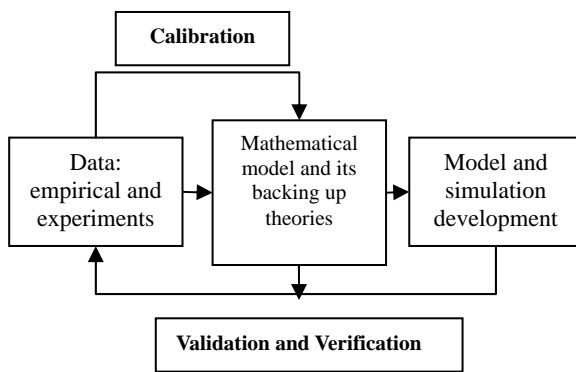


Figure 1. The method used

2.1. Collision Free Multi Agents (Crowd) Motion Planning

A system architecture to deal with a complex VEs has been proposed. The architecture of crowd system allows behaviour of the group and coordination among them to avoid collision [5].

Algorithms that simply co-ordinate agents around a group leader [6] without actually sharing a common goal are faced with a certain number of limitations. For instance, flocking algorithms [3] are vulnerable to local phenomena in the virtual world: if one actor splits from the group it might not be able to reconvene on the basis of the flocking algorithm only, especially if the terrain contains obstacles. More specifically, it is difficult with flocking algorithms to devise strategies for multiple path planning that enable co-ordinate agents to follow different routes to the same destination [7]. In addition to this, flocking method is not easily applicable to specific constraints, such as the maximum dispersion problem. In other words, they go on different routes and reconvene back to a same destination. This problem would generally lead the flock to “loose” some of its elements, as most flocking procedures ensure flock cohesion rather than agent dispersion.

The separate paths that are created for each agent using linear interpolation as implemented in crowd behaviour [8; 9] do not involve a path planning process. The movement is based on guided crowd behaviour from one interest point (IP) to another IP and collision detection with obstacles is

included in separate module for each agent. It is impossible for the crowd behaviour to plan the path and to find the shortest path to destination.

Schooling behaviour [10] particle systems [11], voxel spaces [12] as well as flocking algorithm are ill-suited to the simulation of human movements in planar environment. Their main purpose is to simulate the movement of agents in open space. Although Bouvier et al [11] have successfully used particle system in simulating human crowd movement, it is difficult to include proper path planning in this kind of simulations. Moreover, special purpose actions such as obstacle avoidance are controlled by a separate behaviour routine, for instance avoiding-static-obstacles and avoiding-fish in schooling behaviour.

2.2. Modelling Pilgrimage Movements

We have observed available video materials [See Appendix] and the characteristic features of crowd movement can be summarized as follows: (1) Individuals start pushing, and interactions among people become physical in nature. (2) Crowd flow is further slowed by fallen or injured people acting as ‘obstacles’. (3) Moving and, in particular, passing of a narrowing lane becomes uncoordinated, (4) Jams build up, (5) The physical interactions in the jammed crowd add up (6) People show a tendency towards mass behaviour, that is, to do what other people do [13]. (7) Alternative exits are often overlooked or not efficiently used in escape situations [14].

These observations have encouraged us to model the collective phenomenon of crowd flow in the framework of intelligent self-driven pilgrims. The developed computer simulations of the crowd dynamics of pilgrims are modelled as physical based with explicit visual interaction. The explicit visual interaction might represent a generalized force model [15]) inside the collective behaviour of the pilgrims.

We assume a mixture of socio psychological [16] and physical forces influencing the behaviour in a crowd. It is assumed that each pilgrim is subject to “mixed forces” that represent motivation to move ahead toward the target location. The force here is not meant the real physical force that has dimension of Newton (kg m/second²) but only the analogy of the force that characterizes the internal driving force or motivation of the pedestrian.

Modelling the pilgrims movement starts with assumption that each pilgrim is subject to the force analogy that characterizes the internal moving force or motivation of the pilgrims to move ahead toward the target location.

The force is assumed proportional with the discrepancy between the summation of intended velocities, $\hat{v}(t)$ and the actual current velocity, $v(t)$. Thus, the analogy has same dimension with real force, meter/second².

$$f(t) = m'a(t) = m' \frac{dv(t)}{dt} \quad (1a)$$

$$m' \frac{dv(t)}{dt} = \frac{\sum \hat{v}(t) - v(t)}{\phi} \quad (1b)$$

$$m \frac{dv(t)}{dt} = \sum \hat{v}(t) - v(t) \quad (1c)$$

The intended velocity is a kind of predicted velocity on which way the pilgrim is going to move in the next time ahead. The intended velocity directs the acceleration and the forward force toward the target point. The direction of the intended velocity must be the same as the force and the acceleration. Adopting the Newton law, the acceleration is proportional to the force with a constant proportion called mass, m . Since the acceleration has the same direction as the force, it is also the direction of the acceleration. When the intended velocity is equal to the current velocity, the force (and the acceleration) has zero value and the pedestrian may be stopped or walking with constant velocity.

To simplify the model $m' \frac{dv(t)}{dt} = \frac{\sum \hat{v}(t) - v(t)}{\phi}$,

the parameter mass, m , contains a time dimension and measured in time unit, where the constant ϕ has the time dimension while the mass, m' , is dimensionless. However since both m' and ϕ are constant, they can be put together as one parameter $m = m' \phi$. The parameter mass relates the speed and the acceleration. That means mass has no real physical dimension but obeying the nature of Newton law.

Each pilgrim in the system is influenced by the three forces. Four parameters of the model are the mass, m , alpha, α , beta, β and chi, χ . Alpha influences the force to move ahead, Beta for collision avoidance and Chi for move away. The mass is applied toward the three forces together (global parameter) while the other three parameters are applied only for the particular force. The three forces are

$$\begin{aligned} ma_f &= \frac{\hat{v}_f}{\alpha} - \frac{\alpha'}{\alpha} v \\ ma_a &= \frac{\hat{v}_a}{\chi} - \frac{\chi'}{\chi} v \\ ma_r &= \frac{\hat{v}_r}{\beta} - \frac{\beta'}{\beta} v \end{aligned} \quad (2a, b, c)$$

Resultant of the three forces (addition of vectors) yields

$$m(a_f + a_a + a_r) = \frac{\hat{v}_f}{\alpha} + \frac{\hat{v}_a}{\chi} + \frac{\hat{v}_r}{\beta} - \left(\frac{\alpha'}{\alpha} + \frac{\chi'}{\chi} + \frac{\beta'}{\beta} \right) v \quad (3)$$

Let $a = a_f + a_a + a_r$ and $c = \frac{\alpha'}{\alpha} + \frac{\chi'}{\chi} + \frac{\beta'}{\beta}$

$$ma + cv = \frac{\hat{v}_f}{\alpha} + \frac{\hat{v}_a}{\chi} + \frac{\hat{v}_r}{\beta} \quad (4)$$

Generally, the movement of a pilgrim is from the current location, $\mathbf{p}(t)$ toward the destination point, $\mathbf{e}(t)$. Alpha is applied as the force to move ahead that directs the pilgrim to move. The alpha force makes the pilgrim path almost in a straight line during absence of the other two forces. The direction of the alpha force is from the current location toward the destination. The gradient (direction) of the alpha force is given by $\mathbf{g}(t)$,

$$\mathbf{g}(t) = \frac{\mathbf{e}(t) - \mathbf{p}(t)}{\|\mathbf{e}(t) - \mathbf{p}(t)\|} \quad (5)$$

If there is no obstruction, the pilgrim's intended velocity reaches the maximum walking speed, μ_{max} or smaller ($0 < \hat{v}(t) < \mu_{max}$). The existence of other pilgrims or obstructions will give other forces reduce the walking speed. Thus, the intended velocity for the alpha force is given by

$$\hat{v}_f(t) = \frac{\mu_{max}}{\alpha} \mathbf{g}(t) = \frac{\mu_{max}}{\alpha} \frac{\mathbf{e}(t) - \mathbf{p}(t)}{\|\mathbf{e}(t) - \mathbf{p}(t)\|} \quad (6)$$

The norm in the denominator of equation above represents the distance between the current position and the destination.

When two pilgrims nearly collide, they usually move [repulse] away from each other within a certain distance [Figure 2]. They may not wait until their distance becomes too close to move away unless there are no space surrounding them. A similar behaviour happens when a pilgrim is following another slower pilgrim. They have to adjust their velocity speed so that any proper alignment can be reached matching velocity with nearby pilgrim. Align an actor's velocity vector with that of the local flock.

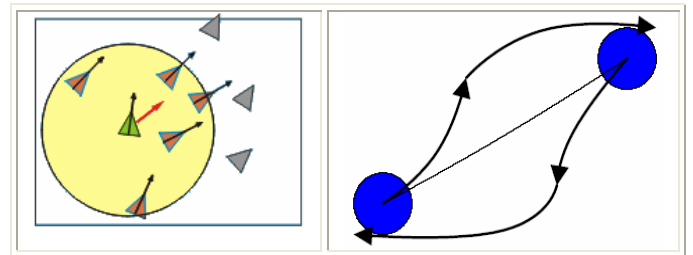


Figure 2. Alignment of Reynold's flock rules and repulse away[3, 17]

If $d(t)$, y and r are, respectively, representing the distance between the pilgrims, interference of the closest pilgrim in the area in front of the actor and the influence radius of pilgrim, the intended velocity of pilgrim i^{th} , $\hat{v}_a^i(t)$ due to Chi force to move away, is given by

$$\hat{v}_a^i(t) = \frac{\mu_{max}(2r - y(t))}{\chi d(t)} = \frac{\mu_{max}(2r - y(t))}{\chi \|\mathbf{p}_k(t) - \mathbf{p}_i(t)\|} \quad (7)$$

By the alpha and chi forces, the pilgrims can adjust the distance between two pilgrims and are able move away from each other. However, those two forces may not be able to prevent a collision when there are many pilgrims in the arena. To more prevent a collision, a force that considers all surrounding pilgrims is needed. For this purpose, it is assumed that each pilgrim has an influence radius that represents his or her security awareness. By giving influence radius they are able to keep a certain distance away from nearest pilgrim and to avoid collisions with nearby pilgrim. The force must be generated when at least two pilgrims' influencing radii partly cover each other as shown in Figure 3.

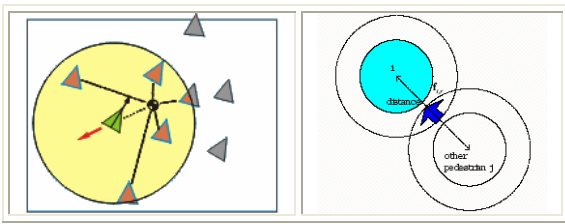


Figure 3. Separation of Reynold's flock rules and Intended velocity to avoid collision by applying beta force [3, 17]

No repulsive force is generated if the influence radius does not overlap each other. Instead of considering the closest pedestrian as the first repulsive force, the second repulsive force considers all surrounding pedestrians and the forces are summed up linearly.

Similar to the first repulsive force, the second repulsive force depends on the distance between the actor and other pedestrians surrounding it. The beta force is given

Since $v(t) = \frac{dp(t)}{dt}$ and $a(t) = \frac{dv(t)}{dt} = \frac{d^2p(t)}{dt^2}$, the formulation can be put together in terms of the current position of pedestrian i , $p_i(t)$ as a second order differential equation

No repulsive force is generated if the influence radius does not overlap each other. Instead of considering the closest pedestrian as the first repulsive force, the second repulsive force considers all surrounding pedestrians and the forces are summed up linearly.

Similar to the first repulsive force, the second repulsive force depends on the distance between the actor and other pedestrians surrounding it. The beta force is given

$$m \frac{d^2 p_i(t)}{dt^2} + \frac{dp_i(t)}{dt} = \mu_{max} \frac{e(t) \cdot p_i(t)}{a \|e(t) \cdot p_i(t)\|} + \frac{2r \cdot y(t)}{\chi \|p_i(t) \cdot p_i(t)\|} + \sum_j \frac{2r}{\|p_i(t) \cdot p_i(t)\|} \cdot \frac{p_j(t) \cdot p_i(t)}{\beta \|p_j(t) \cdot p_i(t)\|} \quad (8)$$

Equation (8) is a non-linear second order differential equation of pedestrian positions that depend on each pedestrian's positions, speeds and accelerations. The analytical solution of the differential equation is very difficult and not practical since it is also dependent on the number of pedestrians and the sight distance. Numerical method through simulation is more favourable and it has the benefit to visualize the movement of each pedestrian in a plan as an animation.

3. Principal of the Model Simulation and Its Results

The differential equation (8) is solved numerically by divide and conquer algorithm using Euler method, which provides adequate results while keeping the computational speed reasonable. Each equation is computed one by one, as each pedestrian is assumed an autonomous agent. A pedestrian has his own internal forces and influence other pedestrians only through his position. The pedestrian movement is based on the resultant forces that act upon him. Other numerical method to solve the differential equation

such as Runge Kutta may produce a better approach to the differential equation but it decreases the computational speed significantly if the number of pedestrian is more than 100, thus, it is recommended for further study.

3.1. Characteristics of the Pilgrim Microscopic Model Developed

In case of microscopic characteristics of pilgrims movement, the effects of forces applied are explained from the Figures 4a, b, c.. It could be seen forward forces cause the paths are linear line.

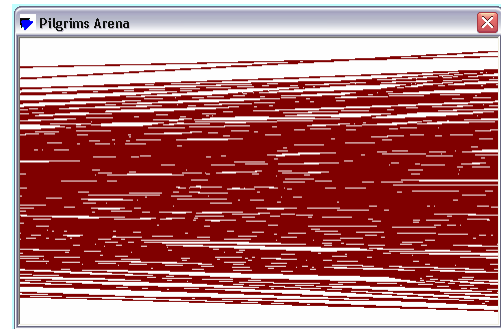


Figure 4a. Forward

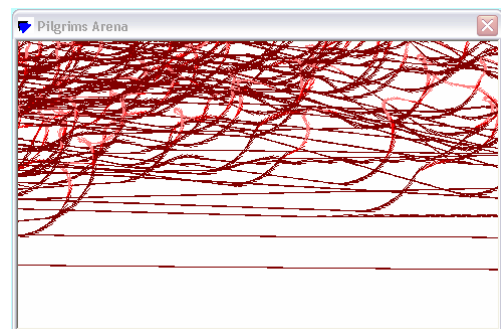


Figure 4b. Forward, and repulse away

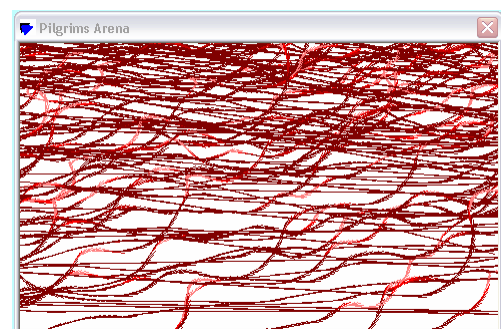


Figure 4c. Forward, repulse away and avoid collision

. It could be seen repulse away forces cause the paths are curving and the forces direct the overtaking behaviour.

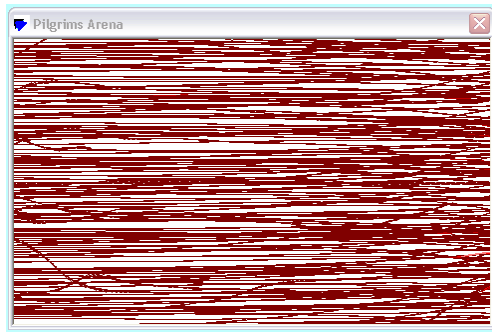


Figure4d. Forward and avoid collision The effects of forces to pilgrims paths

The relationship between speed and density is common to be observed in many traffic and pedestrian model and it is liner. Interestingly, the microscopic simulation model also produces linear graphs but the slope is not consistently negative [Figure 5]. The data is resulted with maximum speed only one and all three forces applied.

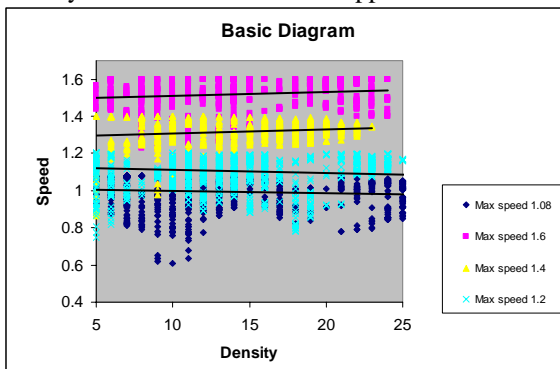


Figure5. The relationship between speed and density

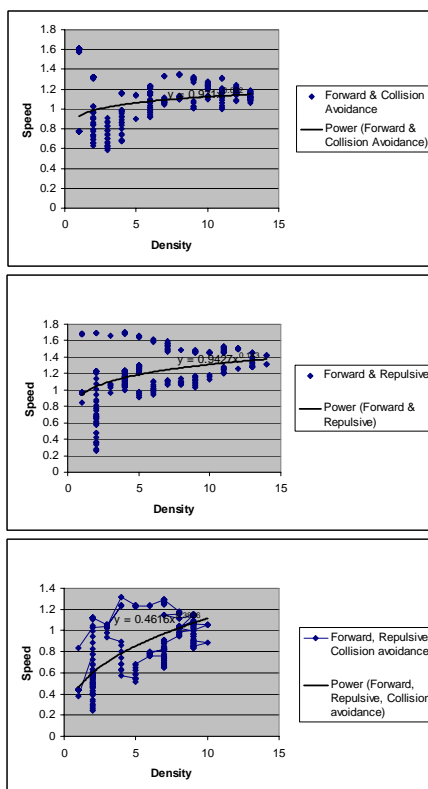


Figure5a, b, c. The relationship between speed and density vary in forces applied and speeds are distributed

normally

Other experiments vary in forces applied and speeds are distributed normally clearly depict positive relationship but will achieve a flat line [Figure 6]. It is interesting to note that the gradient and the intercept of the graph in crowd are different from the traffic [Figure 7].

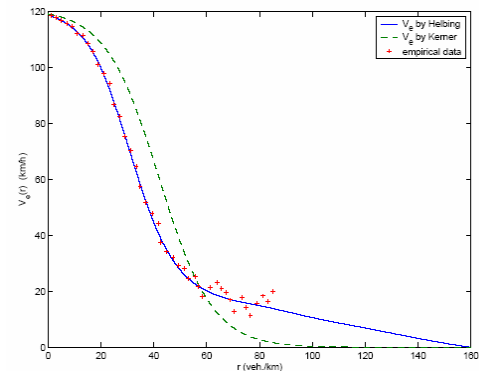


Figure6. The relationship between speed and density by Helbing [18], Kerner[19], traffic data the Dutch highway A9

Compared to similar graph [Figure 7], the speed-density relationship is linear with higher maximum speed on the top of the lower one. The simulation results share this feature.

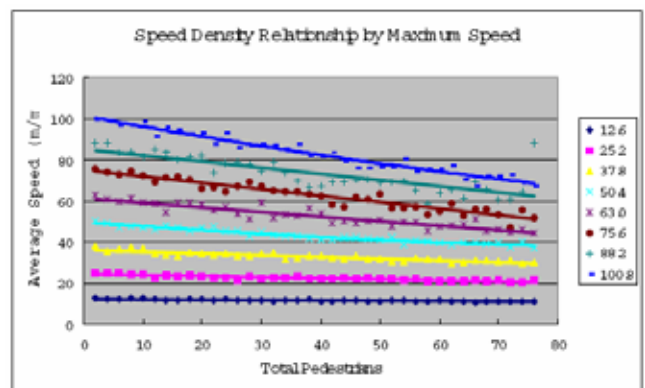


Figure7. The relationship between speed and density [17]

Regarding the validation and verification, crowd compiler will be used to visually compare between real world [Figure 8] and simulation snapshots.





Figure 8. Crowd compiler

4. Conclusion

The microscopic pilgrim simulation model has been developed to determine the microscopic characteristics of pilgrim flow. From the simulation results, there is relationship between density and pilgrim speed. The relationship is influenced by maximum speed and velocity distributions.

Acknowledgments

The study was partially supported by the Research Management Centre, Universiti Teknologi Malaysia. We wish to acknowledge with gratitude the special communicating email with Kardi Teknomo, Tohoku University

References

- [1] Helbing, D., Farkas, I., and Vicsek, T. Simulating Dynamical Features of Escape Panic. *Nature* 2000. Vol.407: 487-490.
- [2] Reynolds, C. Flocks, Herds and Schools: A Distributed Behavioural Model. *Proc. SIGGRAPH'87, Computer Graphics*, v.21, n.4, July, 1987
- [3] Reynolds, C.W., Interaction with Groups of Autonomous Characters. *Proceeding of Game Developers Conference 2000, CMP Game Media Group [formerly: Miller Freeman Game Group], San Francisco, California, 2000: pp. 449-460.*
- [4] Overmars, M. H.. Recent Developments in Motion Planning. *International Conference on Computational Science 2002: pp.3—13*
- [5] Abd. Latiff, M.S., Widyarto. S., "Collision Free Multi Agents Motion Planning in Complex Virtual System", *Proceeding of the 7th International Conference on Work With Computing Systems, WWCS 2004, Kuala Lumpur, June 29- July 2, 2004 H.M. Khalid, M.G. Helander, A.W. Yeo (Editors) . Kuala Lumpur: Damai Sciences, 849-854*
- [6] Brock, D.L., Montana, D.J. & Ceranowicz, A.Z., Coordination and Control of Multiple Autonomous Vehicles. *Proceedings of the IEEE Conference on Robotic and Automation. 1992, Nice, France*
- [7] Abd Latiff, M.S. and Cavazza, M. Synchronous Path Planning for Multi-Agent Co-ordination. *Proceedings of the VSMM 2000 Conference. Gifu, Japan.*
- [8] Thalmann, D., Musse, S.R, Kallmann, M., From Individual Human Agents to Crowds. *Informatik/Informatique, 2000. No1*
- [9] Farenc, N., Musse, S.R., Schweiss, E., Kallmann, M., Aune, O., Boulic, R. and Thalmann, D., 2000. A Paradigm for Controlling Virtual Humans in Urban Environment Simulations. *Applied Artificial Intelligence. 18, 1, pp. 81-103.*
- [10] Tu X and Terzopoulos D, Artificial Fishes: Physics, locomotion, perception, behaviour. *Proc. SIGGRAPH'94, Computer Graphics, Annual Conference Series, volume 28, July-August 1994, ACM Press, New York, pp 43-50*
- [11] Bouvier, E., Cohen, E. and Najman, L., From crowd simulation to airbag deployment: particle systems, a new paradigm of simulation. *Journal of Electronic Imaging 1997. 6[1], 94-107, January*
- [12] Zhang, H. and Wyvill, B., Behavioural Simulation in Voxel Space. *Computer Animation'97, University of Geneva. 1997*
- [13] Keating, J. P. The myth of panic. *Fire J. 57-61, 147 (May 1982*
- [14] Elliott, D. & Smith, D. Football stadia disasters in the United Kingdom: learning from tragedy? *Ind. Environ. Crisis Q. 7(3), 205-229 (1993*
- [15] Helbing, D., Farkas, I. J. & Vicsek, T. Freezing by heating in a driven mesoscopic system. *Phys. Rev. Lett. 84, 1240-1243 (2000*
- [16] Helbing, D. & MulnaÅr, P. Social force model for pedestrian dynamics. *Phys. Rev. E 51, 4282-4286 (1995)*
- [17] Teknomo, Kardi, Application of microscopic pedestrian simulation model, *Transportation Research Part F: Psychology and Behaviour Vol 9 issues 1, January 2006, p. 15-27.*
- [18] Helbing, D., *Verkehrsdynamik, Springer-Verlag, Berlin, 1997*
- [19] Kerner, B.S., and Herrmann, M., Local cluster effect in different traffic flow models, *Physica A, Vol. 255, 1999, pp. 163-188*

Development of Spatio-temporal Query Language for Biodiversity Data

Kamruzzaman, Mohd Taib Wahid, Harihodin Selamat

*Department of Information Systems, Faculty of Computer Science & Information Systems
University Technology of Malaysia, K.B. 791
81310 Skudai, Johor, Malaysia
Tel: (607)-5532422, Fax: (607) 5565044
akzaman@siswa.utm.my, taib@fksm.utm.my, harihodn@itp.utm.my,*

Abstract

Biodiversity data plays an important role in a biological information system. Object relational design for handling taxonomy data involves issues such as modeling the taxonomy data and designing efficient access methods for this model. Spatio-temporal databases deal with spatial objects that change over time (for example, they move or they grow): cars, planes, people, animals, forests, etc. Hence, database systems, in particular, spatial and temporal database systems, and geographical information systems (GIS) need to be extended to handle this kind of information. Relationships between temporal and spatial phenomena there has recently been an increased interest in designing spatio-temporal data models and spatio-temporal databases that deal with geometries changing over time. Research focuses on how to extend database technology to better support spatiotemporal applications and queries. In this paper we concentrated on the query processing requirements for typical biological applications. The main objectives of this paper is to develop spatio-temporal query language and object-relational features of the underlying database management system Oracle to improve the extensibility and flexibility of biodiversity data models, enhance interoperability and analyzing potential as well as to ensure consistency by defining standards based on biodiversity data.

Keywords: Spatio-temporal, Object-relational, Query language, Biodiversity, Data model

1. Introduction

Spatio-temporal data models and query languages have received much attention in the database research community because of their practical importance and the interesting technical challenges they pose. Spatial and temporal databases have long been studied on parallel, but relatively separate tracks. As natural and man-made objects are often associated with both space and time, there has been a growing interest in the integration of spatial and temporal components [1], [2], [3] and [4] leading to the emerging research field of spatiotemporal databases [5], [6].

Since 1970 when Codd [7] proposed the relational model, database system technology has introduced several important changes in the way data are stored and managed. The relational data model has had a great impact on commercial products, mainly because of the development of the SQL language, which is included several additional features with respect to those specified by the theoretical definition of the relational model. Since its initial definition, SQL has been widely extended [8]. A relevant set of extensions has dealt with the introduction of object modeling capabilities [9], resulting in the notion of object-relational data model. Such a model combines the simplicity and the power of SQL the ability of describing new data types with

their associated operations, typical of the object-oriented approach. The object-relational data model is thus a powerful model combining the best aspects of two different approaches.

Spatial data management is often combined with temporal data management, because in many cases one needs to record the temporal evolution of spatially-related entities. The resulting data models are thus termed spatio-temporal data models. One of the most crucial issues when dealing with spatio-temporal databases is the management of the information concerning moving objects in a spatial context. Such an issue represents an important requirement in several application domains, like air traffic control and habitat control of endangered species and so on. So far, the GIS systems have handled the spatial and the non-spatial data separately, which increases the complexity of maintaining data integrity. The use of an object-relational database system to manage this kind of data represents a good alternative, in particular because such a system is able to homogeneously and efficiently manage user-defined information, and to improve integrity for data of different nature.

Most of the spatial-temporal data approaches proposed so far do not exploit the powerful modeling and management features that are provided by recent versions of

commercially available DBMSs. Another main limitation of current approaches is that they do not support multiple granularities in the representation of spatio-temporal data. Multiple granularities, defined as a set of measure units for space and time, are crucial in facilitating the management of information for applications such as air traffic control, meteorological forecast and so forth [10], [11]. The goal of the work reported in this paper is to address such limitations by developing an object-relational approach to the management of spatio-temporal data supporting biodiversity for both space and time.

In particular, in the paper we propose a spatio-temporal query language for biodiversity data expressed in the SQL3 standard. Our start point is the object-relational model and than query language that verifies the model from which we develop an extension of the spatial and non-spatial data types defining a new data type composed the two parts; the first part is the object value and the second the valid time for this value. This extension provides the support required to model multiple spatio-temporal granularities and tools to operate on objects with different granularities in order to address their integration and inter-operability.

2. Related works

Much previous work focuses on either temporal or spatial information, rather than both. For the instance in the temporal domain, interval-based time models [12] were followed by TSQL2's implicit-time model [13] and point-based time models [14]. SQL extensions to express spatial queries were proposed by several authors, including [15] and [16].

The SQL3 standard [17] is the reference language for the object-relational model. It has been defined by extending the previous SQL92 standard with the ability to modify, retrieve and define the data types needed to represent a large variety of application domains. Examples of those extensions include: XML, MULTISSET (like the ARRAY data type, without implicit order), BIGINT and others.

Compared with considerable research on temporal relational databases (e.g. [18], [19], a smaller amount of work has been reported in the context of temporal object databases [20]. This has similarly been the lack of an agreement for a common object model and query language [21]. However, a consensus is now emerging, with the Object Data Management Group (ODMG) effort on standardising Object-Oriented Database (OODB) and a recent release of ODMG 3.0 [22]. The proposed ODMG standard encompasses an object model (ODMG object model), an object definition language (ODL) and an object

query language (OQL). This paper addresses the spatio-temporal data model and an associated query language.

3. Spatio-Temporal Queries

From a design point of view, our intention is not to devise a new spatiotemporal query language from scratch but to appropriately extend the widespread database query language standard SQL. We profit from the fact that the underlying data model rests on the ADT approach which necessitates only conservative extensions to SQL. These are essentially (i) a set of spatio-temporal operations, (ii) temporal selection, (iii) projections to space and time, (iv) spatio-temporal aggregation, and (v) an extension mechanism for the construction of new, more complex spatio-temporal predicates. The benefit of this approach is the preservation of well known SQL concepts, the high-level treatment of spatio-temporal objects, and the easy incorporation of spatio-temporal operations and predicates. Users can ask either standard SQL queries on standard data or use STQL features to inquire about situations involving spatial, temporal, or spatio-temporal data.

The query facility of SQL is provided by the well known SELECT FROM WHERE clause. The integration of predicates like "<" or "<" for standard data types such as integers or strings is well understood. In particular, there are only a few of them which allow one to include them as built-in predicates. When considering more complex and more structured data such as points, lines, or regions, one can try to systematically derive all reasonable predicates. Temporally enhanced object-relational system can express temporal queries as powerful as those of TSQL2 with only minimal extensions to standard SQL by using construct such as user-defined functions and table expressions supported in Object-Relational systems.

We will consider queries from two (simplified) application scenarios. The first scenario is related to biodiversity data (such as flora, fauna) management which pursues the important goal of learning from past biodiversity data and their evolution. We assume a database containing relations with schemas

Flora (FloraName: VARCHAR2, Territory: Species name)
 Flora TypeType (ForestArea: VARCHAR2, Extent: species)
 Area (ForestArea: VARCHAR2, Location: Point)
 Collectors (CollectorName: VARCHAR2, Actor: Person)

The relation flora records the location and the development of different species (attribute Territory) growing and shrinking over time through clearing, cultivation, and destruction processes, for example. The

relation of flora collects from different places and its growth of different data by different people to their extinction (attribute Extent). The relation collectors describes about each flora being on duty from their start at the collection data up to their return (attribute Location).

The second scenario, finally, relates to a database about the forest fire relation in which the evolutions of forest fires are recorded.

ForestFire (FireName: VARCHAR2 (30), Location: Point, StartDate: DateTime, EndDate:DateTime)

3.1 Temporal Selections

The first queries refer to the simple temporal range query. A *temporal selection* extracts the value of an object at a certain instant or the temporal development over a certain period. We can then ask queries like “what happened to the data between days January 10, 1995 to January 20, 1995?”

```
SELECT Flora_ID FROM flora
WHERE (((Flora_collection_year) Between # 01/10/1990#
AND #01/20/1995#));
```

This query shows the functional character of a spatio-temporal object by determining the value of the object at a certain time through a simple function application. A more general version of this query asks what happened to the data between January 10, 1990 and January 20, 1995.

```
WHERE (((Flora_collection_year) Between # 01/10/1990#
AND #01/20/1995#));
```

The “Between” notation specifies a range of time values, that is, a time interval. If a spatio-temporal object is applied to a time interval (or a collection of disjoint time intervals separated by AND), this expression yields a spatio-temporal object restricted to that time interval (function restriction).

3.2 Projections to Space and Time

Projection operations on moving objects map either to their spatial or to their temporal aspect. Assume that we are interested in the geometric locations where the data was changed at the year January 20, 1990. These can be obtained by:

```
SELECT Flora_ID, AreaName
FROM Flora, Area
WHERE ((ForestArea =#Johor# AND #01/20/1990#));
```

This operation computes the *spatial projection* of a spatio-temporal object for the Johor forest area. For an evolving region the trajectory operation returns an object of the spatial type *region* which results from projecting the union of the region values for the Date 20th January, 1990.

The next query asks for the lifespan of a spatio-temporal object: “How many times data have been collected from Johor before 1995?”

```
SELECT COUNT (Flora_ID) AS Ex1,
COLLECTION_DATE AS Ex2
FROM Flora
GROUP BY COLLECTION_DATE
HAVING (COLLECTION_DATE <CONVERT
(DATETIME, '1990-30-12'))
WHERE Area = "Johor"
```

The count operator collects the times when the area Johor is defined (*temporal projection*). In this way inverse temporal functions can be computed. The duration operation computes the length of an interval or of several intervals.

3.3 Aggregations

The following query inquires about the largest collection of flora areas.

```
SELECT Area (max (Extent)) FROM ForestArea WHERE
Type = "Flora"
```

The query demonstrates an example of a *spatio-temporal aggregation* operation max which is an extension of the well known aggregation operator in SQL of the same name. It is here applied to a collection of evolving regions contained in a relation column and computes a new evolving region. Internally, this operator is based on a binary function MAX_{ST} applied to two evolving regions R_1 and R_2 and yielding a new evolving region in the following way:

$$MAX_{ST}(R_1, R_2) := \{(t, r) \mid t \in \text{time} \wedge r = MAX_{geo}(R_1(t), R_2(t))\}$$

This definition uses a function MAX_{geo} which is applied to two regions R_1 and R_2 and which returns larger of both regions.

$$MAX_{geo}(R_1, R_2) = \begin{cases} r_1 & \text{If area}(r_1) > \text{area}(r_2) \\ r_2 & \text{Otherwise} \end{cases}$$

Altogether this means that for 2 evolving regions R_1, \dots, R_n we first compute the evolving region $R = MAX_{st}(R_1, \dots, MAX_{st}(R_{n-1}, R_n) \dots)$. Afterwards, we apply the raise

area of R, which computes the area of R at all times as a temporal real number.

4. Conclusion

In this paper we have discussed issue of spatio-temporal query language. Based on an ADT approach to the integration of spatio-temporal data types into data models we have shown how to extend SQL to a spatio-temporal query language called STQL. We have shown how to obtain query operators by expressing temporal selections and aggregations.

For the first experimental investigation purpose, we have tried to applied johor forest data for the suitability of the model for handling spatio-temporal data. For this purposes, we inserted and manipulated the biodiversity data through proposed model and perform a specific set of pre-defined temporal and spatio-temporal queries on it as discussed above. The proposed model supports well the spatio-temporal data to perform spatio-temporal queries.

Experimental works still under process to obtain better result than previous experiments. Few more steps to be done within few months such as comparison between proposed data model and existing model, user interface design and system development.

Acknowledgements

I would like to thank to my supervisors and friends who are working fsksm databases lab to support me to write this paper as well as my research work. I also would like to thank to sps and ppi for organizing an international conference.

5. Reference:

- [1]. Langran G. 1992. Time in Geographic Information Systems. Taylor & Francis, London Open GIS Consortium Inc.
- [2]. M.F. Worboys. A Uni_ed Model for Spatial and Temporal Information.
- [3]. *ComputerJournal*. 1994. Vol.37, No.1, pp.26-34.
- [4]. Claramunt C, Thériault M. 1995. Managing time in GIS: an event-oriented approach. In: Clifford J, Tuzhilin A (eds) Recent Advances in Temporal Databases. Springer-Verlag.
- [5]. Peuquet D, Duan N. 1995. "An event-based spatio-temporal data model (ESTDM) for temporal analysis of geographical data", *International Journal of GIS* 9: 7-24.
- [6]. Gueting RH, Böhlen *et al.* 2000. A foundation for representing and querying moving objects. *ACM Transactions on Database Systems* 25: 1-42.
- [7]. Spery L, *et al.* 2001. A spatio-temporal model for lineage metadata. *Geoinformatica* 5:51-70.
- [8]. Codd, E. F. 1970. A Relational Model of Data for Large Shared Data Banks. *CACM* 13 (6).
- [9]. Date, C. J. 2003. An Introduction to Database Systems. 8th Edition. Addison-Wesley, Reading, Mass.
- [10]. OMG. 2000. Unified Modelling Language Specification, Version 1-3. Object Management Group.
- [11]. T. Bittner and B. Smith. 2001. "A Unified Theory of Granularity, Vagueness and approximation", In Proc. of COSIT Workshop on Spatial Vagueness, Uncertain and Granularity.
- [12]. V. Katri, S. Ram, R.T. 2002. Snodgrass and G. O'Brien. Supporting User Defined Granularities and Indeterminacy in a Spatiotemporal Conceptual Model. *Special Issues of Annals of Mathematics and Artificial Intelligence on Spatial and Temporal Granularity*, 36(1-2). 195-232.
- [13]. N.A. Lorentzos and Y.G. 1997. Mitsopoulos. SQL Extension for Interval Data. In *IEEE Transactions on Knowledge and Data Engineering*, Vol.9, No.3, pp.480-499.
- [14]. R.T. Snodgrass, et al. 1995. *The TSQL2 Temporal Query Language*, Kluwer.
- [15]. D. Toman. 1996. Point vs. Interval-based Query Languages for Temporal Databases. In *Proceedings of the 15th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pp.58-67.
- [16]. M.J. Egenhofer. 1994. Spatial SQL: A Query and Presentation Language. In *IEEE Transactions on Knowledge and Data Engineering* Vol.6. No.1. pp.86-95.
- [17]. R.H. Gutting and M. Schneider. 1995. Realm-Based Spatial Data Types: The ROSE Algebra. In *VLDB Journal*, Vol.4, No.2, pp.243-286.
- [18]. Eisenberg, A., *et al.* 2004. Spatial Query language (SQL): 2003 has been published. *SIGMOD Record*, vol. 33, no. 1.

- [19]. Tansel, et al. 1993. "Temporal Databases: Theory, Design, and Implementation", The Benjamin/Cummings Publishing Company
- [20]. Snodgrass R. 1995. "The TSQL2 Temporal Query Language. Kluwer Academic Publishers.
- [21]. Snodgrass R 1995. Temporal object-oriented databases: a critical comparison. In Kim W (ed) Modern Database Systems: The Object Model, Interoperability, and Beyond. Addison-Wesley, pp 386-408
- [22]. Huang B, *et al.* 2000. A spatio- temporal object model and query language (extended abstract). In: GIScience 2000: First International Conference on Geographic Information Science. 28-31 October, Savannah, Georgia, USA
- [23]. Cattell R and Barry D. 2000. The Object Data Standard: ODMG 3.0. San Francisco, Morgan Kaufmann Publishers Inc.

Local Position System of a Mobile Robot: a Practical Perspective.

Iman H. Kartowisastro

Computer Engineering Department,
Bina Nusantara University, Jakarta, Indonesia.
Tel : 62-21-534 5830, Fax : 62-21-530 0244; E-mail : imanhk@binus.ac.id

Abstract

Localization problem in determining position and orientation of an autonomous robot with respect to a world reference frame is commonly encountered. Personal service robot which has the largest share in the growth of service robot needs a simple and affordable system to overcome this problem. This paper presents an alternative approach to solve localization problem by utilizing Local Positioning System based on ultrasonic sensors. With this system, 3D information in regards to position and information of a mobile robot can be obtained fast enough to cope with a robot moving with a speed of 30 cm/s. The use of LPS to enhance the performance of control system has been already proven for a mobile robot moving on area with high irregularity surface.

Keywords

Local Positioning System (LPS), personal service robot, absolute measurement, relative measurement, feedback control system.

1. INTRODUCTION

Penetration of service robots into market grows significantly in which estimated growth of market value in 2007 will be 2.5 times of value in 2003 and the population number in 2007 is predicted to be 6 times of population in 2003 [1]. The biggest portion of this growth falls into a category of personal/domestic service robots. Hence, research in this category of service robots are conducted intensively world wide. Furthermore, in the ubiquitous computing environment, service robot, such as IDRO [5], can take advantages of it and this in turn will accelerate the opening of the service robot market in the post PC era.

Works in the field of robot navigation is important as mobile robot platforms which constitute most personal robots depend heavily on the navigation capabilities and environmental navigation in turn provides a basis to carry out tasks. Localization problem in determining position of a robot can be solved via a hybrid approach, namely by using data base of reference images throughout the regions and the use 802.11b wireless network signals [6]. In this

work, directing the camera to obtain a view of the upper portion of the region and polling the wireless signals coming out from several access points will allow computer to process those data in obtaining information about where the robot is. Some researchers proposed a multi agent based architecture for outdoor mobile robot navigation [2]. A mobile robot equipped with CCD cameras, binocular CCD cameras, two LADARs and a GPS device was used in this work, and it illustrated that the architecture improved the reasoning ability about the world by making use of apriori global knowledge. Others work more on algorithm aspects, such as localization algorithm [3]. Simulation study showed that the algorithm gave results in 5.51 seconds with an error of 15 cm. This result is somewhat longer in real time condition, considering during this period a robot may already be moving to somewhere else or may hit anything surrounding it. Localization problem is further expanded into a team of robots in which each robot is equipped with proprioceptive and exteroceptive sensors [7], [8]. An extended Kalman filter for fusing the data coming out from these 2 sensors are used and from simulation result the accuracy on the localization is improved by the use of relative bearing.

Even though there are so many research carried out in the field of localization and navigation, however, few works

are dedicated to producing simple systems but robust enough which are developed specifically for personal service robots in which price, affordability, and simplicity aspects are very important. There are so many algorithms developed in localization subjects, but implementations are sometimes not easy due to obtaining data in real time fashion and impractical problem imposed. This paper presents a practical perspective of developing a Local Positioning System based robot control system which in turn will enhance solution for navigation problem. In our earlier studies, we investigated the use of ultrasonic sensors in knowing the location of an object [4]. The system was then further developed to acquire a Local Positioning System specifically developed for personal service robot.

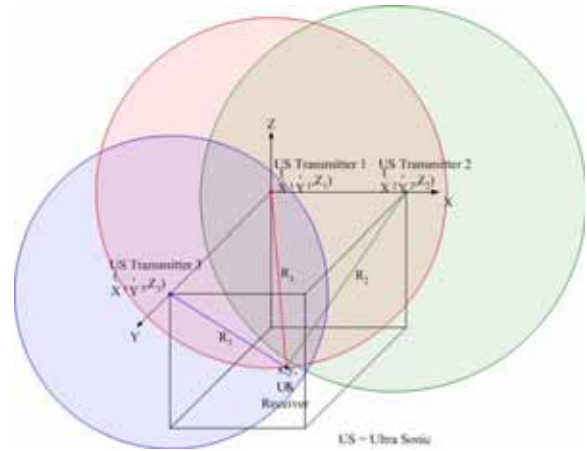


Figure 1.
Ultrasonic Based Local Positioning System (LPS)

2. SYSTEM DESIGN AND IMPLEMENTATION

The design of robot control system is basically trying to have a mobile robot moves in a such way so that the desired position and orientation can be achieved without sacrificing system stability. In the real condition, robot usually moves on an uneven floor, and not on the one with smooth surface. Apart from developing better control algorithm, some other approach needs also be devised. Local Positioning System based control system with ultrasonic signal as a complement to a conventional control system is proposed by still keeping the price low.

Basically, the system consists of 2 main parts, namely :

1. Local Positioning System (LPS) to allow a 3 dimensional tracking of robot movement.
2. A differential system based wheeled mobile robot with proportional control algorithm.

LPS is built using 2 ultrasonic receivers mounted on the object to be monitored (mobile robot) and 3 ultrasonic transmitter attached on the corners of a ceiling as shown in figure 1. The distance between transmitter 1, transmitter 2, transmitter 3 and receiver are measured as R1, R2 and R3, respectively. The use of 2 receivers (front and rear) on the robot allows the system to know the orientation of the robot.

With the use of a set of trigonometric equations [4], the absolute position (X,Y,Z) of the 2 ultrasonic receivers can be obtained as

$$x = \frac{R_1^2 - R_2^2 + X_2^2}{2 X_2} \quad (1)$$

$$y = \frac{R_1^2 - R_3^2 + Y_3^2}{2 Y_3} \quad (2)$$

$$z = \pm \sqrt{R_1^2 - (x^2 + y^2)} \quad (3)$$

As positions of 2 points in the robot are obtained, the orientation can also be found accordingly. More detailed about analytical works of this localization approach can be found in this reference.

A tone detector circuit is used to recognize the ultrasonic signal. Communications among these ultrasonic receivers and transmitters is controlled by a microcontroller (Atmel 89C52). The distance of each receiver on the object from all 3 transmitters is obtained by measuring the time elapsed and both receivers communicate with PC using Infra Red (IR) signal.

The block diagram of this system is shown in figure 2.

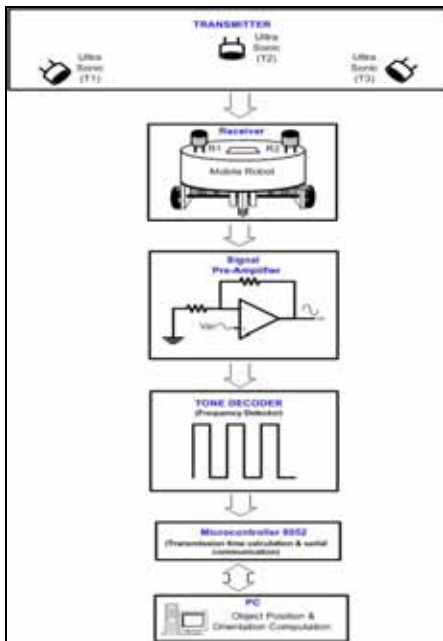


Figure 2. Block Diagram of LPS.

Communication between ultra sonic transmitter and receiver is executed in a half duplex fashion using protocol as shown in figure 3.

{	1	}	Receiver asks Transmitter 1 to send signal
Start Byte	Transmitter id	Stop Byte	
{	2	}	Receiver asks Transmitter 2 to send signal
Start Byte	Transmitter id	Stop Byte	
{	3	}	Receiver asks Transmitter 3 to send signal
Start Byte	Transmitter id	Stop Byte	
{	M	}	Acknowledgement signal from transmitter to receiver
Start Byte	ACK	Stop Byte	

Figure 3 Protocol of ultrasonic transmitter and receiver.

Once the elapsed time between receivers and transmitters has been processed by the microcontroller, then these values are sent to PC using Infra Red (IR) signal. PC then computes these data to obtain the position and orientation of the robot. The protocol of this communication is depicted in figure 4.

{	A	TH2	TL2	TH1	TL1	}
Start Byte	US 1	U Time Receiver 2	U Time Receiver 2	U Time Receiver 1	U Time Receiver 1	Stop Byte
Receiver unit sends data of elapsed time between transmitter 1 and both receivers to PC						
{	B	TH2	TL2	TH1	TL1	}
Start Byte	US 2	U Time Receiver 2	U Time Receiver 2	U Time Receiver 1	U Time Receiver 1	Stop Byte
Receiver unit sends data of elapsed time between transmitter 2 and both receivers to PC						
{	C	TH2	TL2	TH1	TL1	}
Start Byte	US 3	US Receiver 2	US Receiver 2	US Receiver 1	US Receiver 1	Stop Byte
Receiver unit sends data of elapsed time between transmitter 3 and both receivers to PC						

Figure 4. Protocol between LPS and PC using Infra Red signal.

A complete picture of the whole process in the developed Local Positioning System is given in figure 5.

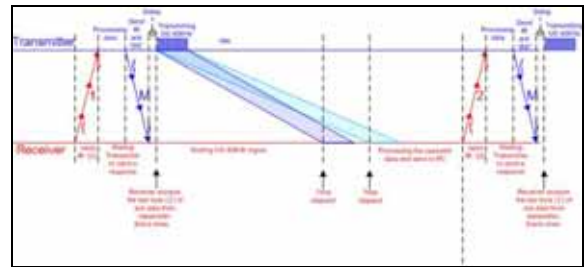


Figure 5 The whole process in the Local Positioning System

Although the receiver is capable of receiving ultrasonic signal from a distance over 4 metres (long enough for a personal robot use), due to limited space reason in our lab, the system is setup to form a workspace of 2m x 2m x 2m. Some experiments were conducted to obtain the characteristics of the system and it was found that the tracking resolution was 3 cm and X,Y,Z measurement rate was achieved at 300 ms using Pentium 3 based PC. The system can track robot with a speed of 30 cm/s comfortably as shown in figure 6. Increasing the robot speed will create errors in the resulting X,Y,Z measurements. This is due to the fact that the measurement of elapsed time is done sequentially from transmitter 1 to transmitter 3. Hence, increasing the speed will cause the distance is no longer measured with respect to 1 single point, instead it is measured with respect to 3 separate points. Hence, the higher the robot speed, the bigger error it creates.

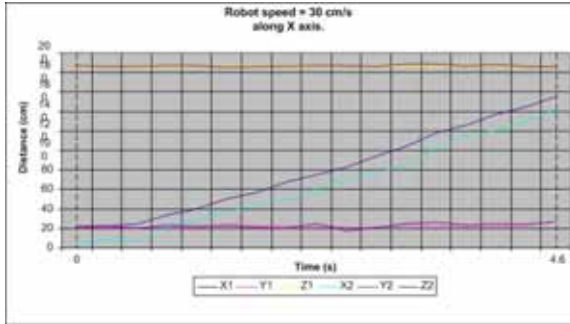


Figure 6
Tracking of a robot movement.

The LPS is then fitted to a wheeled mobile robot with a differential drive system. A trajectory is generated between the initial and final positions of robot trip. The relative position measurement through wheel angular distance is carried along the trajectory while the absolute measurements are taken at the initial and final positions (cartesian space). Hence, the robot has a 2 level closed loop control system and the control algorithm is derived from kinematics information accordingly. As two points of measurement are attached to the robot, this will allow the system to have information about robot position (centre of robot) and orientation as shown in figure 7.

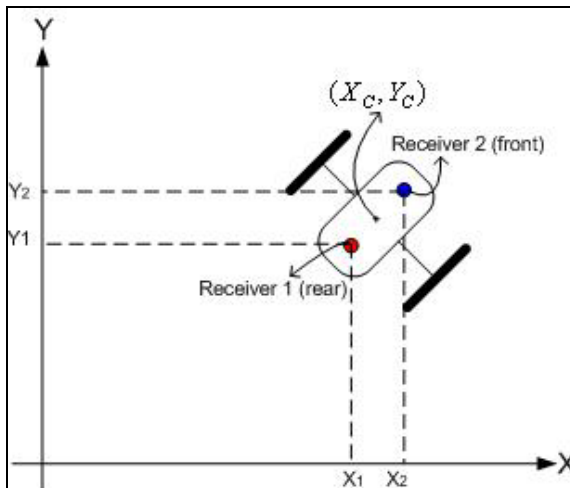


Figure 7.
Measurement of mobile robot position

During implementation, inner loop control system to obtain the desired relative position is executed by AVR microcontroller ATmega8535 whilst the outer loop control system to obtain the desired absolute position is carried out by a PC. The feedback control system diagram is shown in figure 8.

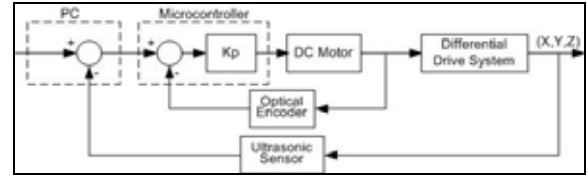
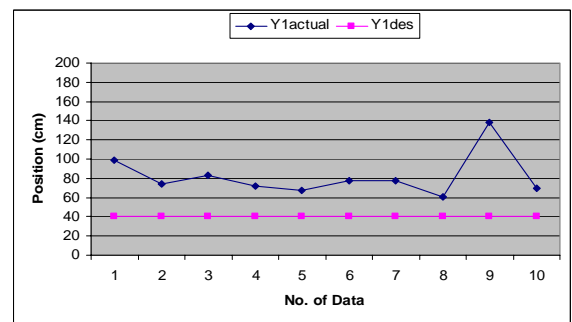
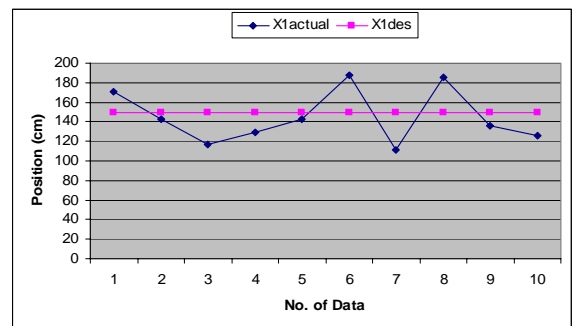


Figure 8.
A Two Level Feedback Control System.

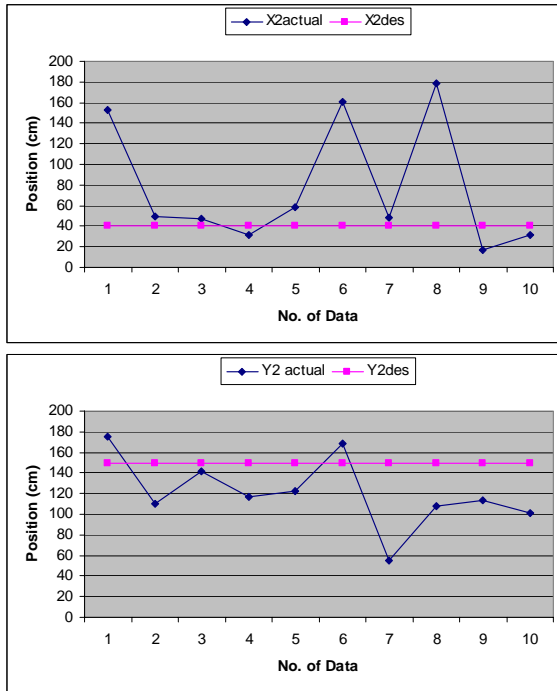
To improve accuracy during each absolute position measurement by LPS, several data can be taken and median filter process is implemented to reduce spike problems in the measurement.

3. EXPERIMENTAL RESULT

The mobile robot equipped with an LPS is located in an area which has irregularities of terrain. The area is setup in such away so that a robot can slide due to the irregularities of terrain. The robot is then given a task to go from an initial position to a desired of 2 consecutive points. An assessment of LPS effectiveness to improve control system is carried out by executing this task with and without feedback signal from LPS. Initial point is set to a coordinate of (150 cm, 150 cm) and the desired 2 consecutive points are set to coordinates of (150 cm, 40 cm), and (40 cm, 150 cm), respectively. Figure 9 and 10 show the results of this experiment with 10 data samples.

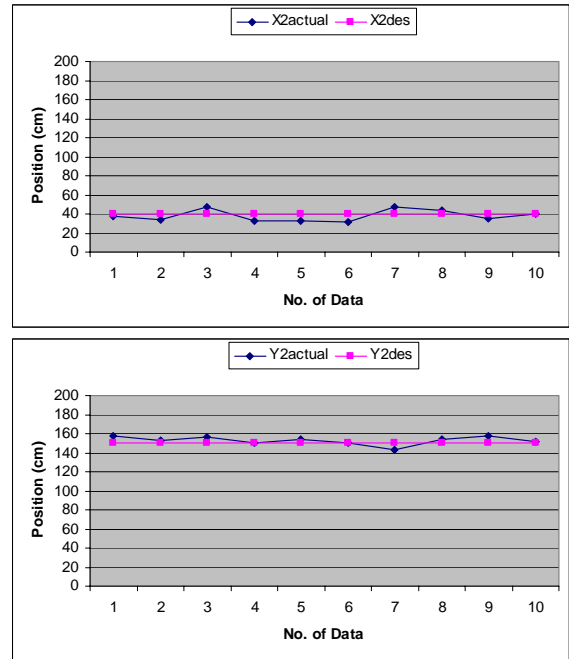


(a)
Response to the 1st Desired Position



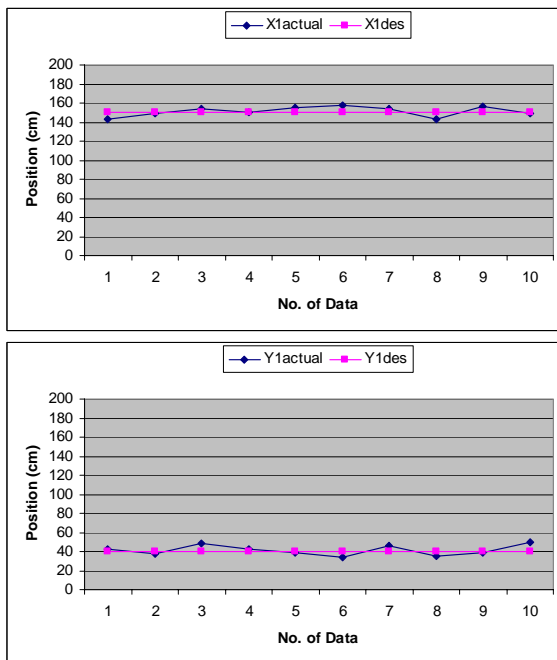
(b) Response to the 2nd Desired Position

Figure 9
Performance of Control System with no LPS Feedback Signal



(b) Response to the 2nd Desired Position

Figure 10
Performance of Control System with LPS Feedback Signal



(a) Response to the 1st Desired Position

Figure 9 indicates that the use of optical encoder to provide feedback on angular position of each wheel (relative measurement) is not effective as highly irregularity of surface will produce non systematic errors. The control system ensures the robot goes to the desired location by checking rotations of the robot wheels. However, due to the possibility of sliding during the movement, the robot does not arrive at the desired location correctly. Errors in the first destination will accumulate to the second destination. The experiment shows that the errors can be as high as 140 cm.

On the other hand, providing absolute position through LPS feedback system can overcome this problem as shown in figure 10. The 2nd level of feedback control system will make correction if the 1st level fails to satisfy in targeting to the desired 1st and 2nd positions. Irregularities due to existence of small stones and gravel during the experiments could not fool the system.

4. CONCLUSION

The approach proposed in this paper about mobile robot localization through the use of Local Positioning System is merely an alternative to solve the problems. Experiments shows that,

1. An inexpensive Local Positioning System consisting of 3 ultra sonic transmitters and 2 receivers provides a modest system for tracking a robot with a speed of 30 cm/s.
2. LPS based control system is very effective for a wheeled mobile robot operating on an area where non systematic errors occur due to high irregularity of terrain.
3. The use of PC for a 2nd level feedback control system allows flexibility in developing more advanced control algorithm.

Currently, for further development, a work is being conducted to expand the operating area by using a roaming technique from 1 workspace into another workspace.

REFERENCES

- [1] United Nations-Economic Commission for Europe and International Federation of Robotics, "World Robotics 2004 : Statistics, Market Analysis, Forecasts, Case Studies and Profitability of Robot Investments.", 2004.
- [2] H. Zhang; M. Zhu, ' Self Organized Architecture for Outdoor Mobile Robot Navigation", Journal of Zhejiang University, ISSN 1009-3095, pp.583-590.: <http://www.zju.edu.cn/jzus/2005/A0506/A050618.pdf> (retrieved in January 2006)
- [3] T. Bilgic and IB Turksen, "Model Based Localization for an Autonomous Mobile Robot Equipped with Sonar Sensors. : http://www.ie.boun.edu.tr/~taner/publications/papers/ieeesm_c95.pdf (retrieved in January 2006)
- [4] IH. Kartowisastro, A. Suhartono, Agus;Sugihyanto;S. Sumendap, "Local Positioning System in A Limited Room Space", Research Memo, Computer Engineering Dept, Bina Nusantara University, Jakarta-Indonesia, March 2004.
- [5] Yong K. Hwang, Manjai Lee and Dongman Lee. "Robot's Role in Ubiquitous Computing Household Environments." Proceeding of 35th International Symposium on Robotics, March 2004, Paris-France.
- [6] Paul Blaer and Peter K. Allen. "A Hybrid Approach to Topological Mobile Robot Localization". Proceeding of 35th International Symposium on Robotics, March 2004, Paris-France.
- [7] Agostino Martinelli and Roland Siegwart. "Observability Analysis for Mobile Robot Localization". Proceeding of 35th International Symposium on Robotics, March 2004, Paris-France.
- [8] Agostino Martinelli, Frederic Pont and Roland Siegwart. "Multi-Robot Localization Using Relative Observations". Proceeding of the 2005 IEEE International Conference on Robotics & Automation, Barcelona, Spain, April 2005.

Strategic Information Systems Planning (SISP) and Strategic Information Security Planning (SISecP) In Malaysian Government Agencies

Md Hafiz Selamat^a, Mohd Adam Suhaimi^b, Husnayati Hussin^c

^aFaculty of Computer Science and Information Systems
Universiti Teknologi Malaysia, Skudai, Johor.
Tel: +607-5532423, Fax: +607-5565044, E-mail: hafiz@fksm.utm.my

^bKulliyah of Information Communication Technology
International Islamic Univeristy Malaysia, Gombak, Selangor
Tel: +603 - 61965601, Fax: +603 - 61965179, E-mail: ^badam@iiu.edu.my; ^chusnayati@iiu.edu.my

Abstract

Strategic information systems and strategic information security are two different attributes in information systems world. Information systems security must be integrated with business plan[6] and strategic information systems planning must be align with business plan. This study aims to investigate the contribution of information security attributes to the strategic information systems planning in the organization especially to the Malaysian Government Agencies. Strategic information system planning (SISP) is an exercise or ongoing activities that enable organization to develop priorities for information system (IS) development[1]. A completes SISP approach is a combination of methods, processes and implementation[2]. As a new business strategies and information technologies are both rapidly moving targets, it is a very challenging task to produce an effective plan that achieves business objectives with efficient information systems support[3]. Organization invest very large amount of time and money in the SISP project. In a typical SISP project, term of key managers, users, selected clients, and IS specialist are formed[3] and planning methodology is chosen. On the other hand, Information security planning is to mitigate risk associated with the processing of information with confidentiality, integrity and authenticity[6] with ten main domain areas as common body of knowledge in information security. Finally, this study will explore both attribute of SISP and SISecP usage in information systems planning.

Keywords: strategic information systems planning, strategic information security; information systems security

1.0 ICT Strategic Plan Overview

Information and Communication Technology (ICT) is an increasing powerful tool for improving the delivery of the government service, enhancing ICT development opportunities and participating in global marketplace. A strategy is the pattern of missions, objectives, policies, and significant resource utilization plans stated in such a way as to define what business the agency is in (or is to be in) and the kind of agency it is or is to be, therefore strategy and business visions are closely linked. A complete statement of strategy will define:

- Service line and the customers;
- Customers segments for which services are to be designed;
- Channels through which these customers will be reached;
- The means by which the operation is to be financed;
- Service objectives; and
- The "image" which the agency will project to its customers.

Planning for information systems, as for any other system, begins with the identification of organizational needs. In order to be effective, development of any type of computer-based system should be a response to need-whether at the transaction processing level or at the more complex information and support systems levels. Such planning for information systems is much like strategic planning in management.

Objectives, priorities, and authorization for information systems projects need to be formalized. The systems development plan should identify specific projects slated for the future, priorities for each project and for resources, general procedures, and constraints for each application area. The plan must be specific enough to enable understanding of each application and to know where it stands in the order of development. Also the plan should be flexible so that priorities can be adjusted if necessary. King [5] in his recent article has argued, strategic capability architecture - a flexible and continuously improving infrastructure of organizational capabilities - is the primary basis for a company's sustainable competitive advantage. He has emphasized

the need for continuously updating and improving the strategic capabilities architecture.

Relationship between information system functions and corporate strategy was not of much interest to Top Management of firms. Information Systems were thought to be synonymous with corporate data processing and treated as some back-room operation in support of day-to-day mundane tasks. In the 80's and 90's, however, there has been a growing realization of the need to make information systems of strategic importance to an organization. Consequently, strategic information systems planning (SISP) is a critical issue. In many industry surveys, improved SISP is often mentioned as the most serious challenge facing IS managers.

Organizations nowadays depend largely on computer based Information Systems (IS) for a vital part of their operation. IS comprise the information that is being stored, or in any way processed by an organization, the hardware and software that constitutes the configuration of computer systems, a social system that is formed by the actions and relations among the IS users, as well as a set of procedures that guide the users' actions. Under this perspective, IS have not only a technical part, but also a social dimension. IS are of high significance to organizations across a wide range of economic sectors. In consequence, their proper function and unobstructed operation is a critical issue that has attracted the attention of both IS research and practice. Information systems security management is a stream of management activities that aim to protect the IS and create a framework within which the IS operates as expected by the organization.

2.0 What is Strategic Information Systems Planning (SISP)?

IS/IT planning is the organized planning of IT infrastructure and applications portfolios done at various level of organization. It is very important to both user and planner because end user do IS/IT planning for their own units; end user must participate in the corporate planning, therefore they must understand the process; corporate planning determines how the IS/IT look like and determine what applications end user can deploy. Finally the future of every unit in the organization will be impacted by the IS/IT infrastructure.

SISP is the process of identifying a portfolio of computer-based applications to serve an organization best. Ideally, the organization makes SISP part of its strategic business planning process to link the resulting IT strategy to the business strategy. SISP also includes the specification of databases and systems to support those applications. SISP may encompass the selection of conventional applications that best fill the organization's current and future needs. SISP also may entail the search for new applications with the potential to create an advantage over competitors.

To perform SISP, an organization usually carries out a major, intensive study. Most follow one of several, similar well-defined and documented methodologies; a few spend the time on their own versions. Committees of users and IS specialists are more common, often using the methodology's vendor for training and guidance. During the multi-steps study, a portfolio of applications is defined, along with appropriate priorities, databases, data elements, and a network of computers and communications equipment to support them. The study also provides a schedule for their development and installation. The organization periodically updates the plan after its initial approval.

SISP basically address four general issues; aligning IS/IT plan with the organizational business plan; designing IS/IT architecture for organization in such a way that user, applications and databases can be integrated and network together; efficiently allocating information systems development and operational resource among competing applications and finally planning information project in order to complete on time and within budget and include the specific functionalities.

3.0 The Perspective of Strategic Information Systems Planning

In order to put the planning for strategic information systems in perspective, the evolution of information systems according to the three-era model[7] is pertinent. According to this model there are three distinct, albeit overlapping, eras of information systems, dating back to the 60's. The relationship over time of the three eras of information systems is shown in table 1 below:

Table 1: The Three Era Model of Information Systems[7]

	ERA	CHARACTERISTICS
60s	Data Processing (DP)	Standalone computers, remote from users, cost reduction function.
70s & 80s	Management Information Systems (MIS)	Distributed process, interconnected, regulated by management service, supporting the business, user driven.
80s & 90s	Strategic Information Systems (SIS)	Networked, integrated systems, available and supportive to users, relate to business strategy, enable the business - business driven.

Applications in the overall Data Processing (DP), Management Information Systems (MIS) and Strategic Information Systems (SIS) area need to be planned and managed according to their existing and future contribution to the business. Traditional portfolio models consider the relationship of systems to each other and the tasks being performed rather than the relationship with business success.

4.0 SISP Methodology

The task of strategic information systems planning is difficult and often organizations didn't know how to do it. Strategic information systems planning is a major change for most organizations, from planning for information systems based on users' demands to those based on business strategies. Strategic information systems planning can change the planning characteristics in major ways. For example, the time horizon for planning changes from 1 year to 3 years or more and development plans are driven by current and future business needs rather than incremental user needs. Increase in the time horizon is a factor which results in poor response from the top management to the strategic information systems planning process as it is difficult to hold their attention for such a long period.

SISP methodology is a standard set of techniques and supporting tools to study the organization to produce SISP[8]. Generally, several methodologies were applied for SISP success[9]. The SISP team normally uses methodologies as a guideline in the SISP process. To accomplish SISP, the organization should identify the application portfolio, databases definition elements of data network and communication to support the applications. So, the SISP is a complex task [10]. Preparing the schedule for study, analysis, development and installation of information systems also required appropriate methodologies.

Methodologies for application is depends on the scope of the study and it can be dragged from several week to several months[9]. There are various types of methodologies and the SISP team should select to support their purposes. Most of the methodologies are owned and referred to the designer or consultant of the methodology [11][12]. Some vendor such as Anderson Consulting, Pro planner, Price Waterhouse Copper and many others provide SISP methodologies integrated with the software package in CASE tool. This tool need standard environment, common approach, standard vocabulary and standard user interface. CASE tools also provide communications and integration and then let various user works simultaneously on different part of study [9][12]. The previous review found the SISP methodology indicates five main components to produce SISP Plan as figure 2.0. They are project initiation, organizational analysis, identifying IT strategy, identifying the benefit from the use of IT and finally, designing the implementation plan[12]

4.1 The Strategic IS/IT Planning Process

The planning process in strategic IS/IT planning consist input, output and processing activities[7]. Figure 3.0 shows the input and outputs of the planning process. The input activities are internal business environment, external business environment, internal IS/IT environment, and external IS/IT environment. Internal business environment is the current business strategy, objective, resources, processes and the culture and values of the business. External business environment are the economic, industrial and competitive climate in which the organization operates. Internal IS/IT environment is the current IS/IT perspective in the business, its maturity, business coverage and contribution, skills, resources and technological infrastructure. The current application portfolio of existing systems and systems under development, or budgeted but not yet under way also part of the internal IS/IT environment. External IS/IT environment is the technology trend and opportunities and the use made of IS/IT by external bodies.

The output activities are IS/IT management strategy, business IS strategy and IT strategy. IS/IT management strategy is the common elements of the strategy that apply throughout the organization, ensuring consistent policies where needed. Business strategy is how each unit or function will deploy IS/IT in achieving its business objectives. Alongside each of these are application portfolios to developed for the business unit and business models describing the information architecture of each unit. The portfolios may include how IS/IT will be used at some future date, to help the units to achieve their objective. IT strategy is a policy and strategies for the management of technology and specialist resources.

4.3 Strategic Information System Planning in Malaysia

There are a few SISP researchers in Malaysia and the first researcher is Isa[4] and is about SISP in the Islamic perspective[12]. He explore and introduce new values of SISP which never revel before such as a concept of tawheed (unity), khalifah (trusteeship), worship, ilm (knowledge), halal (praiseworthy), haram (blameworthy), adl (social justice) zulm (tyranny), istislah (public interest) dhiya (waste). Isa also explore the practice of SISP among Islamic Organization in Malaysia such as Jabatan Kemajuan Islam Malaysia, Lembaga Urusan Tabung Haji and Institute Teknologi Mara (now Universiti Teknologi Mara). Another researchers are Zurina[13] and Maslinda [13]. Their researchs are SISP team satisfaction during the completion of SISP and SISP practice selected private sector in Malaysia. Experience in SISP, knowledge and team working are contribute to SISP satisfaction[14][12] and most of private sector in Malaysia have SISP and they are aware the important of SISP to improve organizational activities[13][12]. They

also discover the formation of information systems department in their organization contribute to the success of SISP process. Selamat [15], is using the framework produce by Ward [7] in the secondary school resource environment as a case study. There is a different between Institute of Higher Learning (IHL) business compare to business organization in their SISP framework [16]. The SISP study Malaysia's IHLs is not in depth compare to the IHL in US and Europe. This study need to be highlighted and enhance in order to increase their understanding and SISP implementation in the organization. Suraya[17] was developed a prototype SISP tools which combining combine and automate several planning techniques in a single system and the output is information architecture and also will be the input for the system prototype development.

4.4 SISP in Malaysian Government Agencies

In the context of Malaysian strategic management practice, the model as shown in figure 4.0 is used as a basic framework. The strategic management is link to strategic management process[17]. There are four main component s in the framework; vision 2020 for Malaysian, present and future challenge, process of transformation, and strategic management (include corporate strategic planning and ICT strategic planning). Three ingredients to complementary in a strategic planning process are strategic thinking, strategic intent and strategic action. Thinking is needed to ensure the acquisition of right knowledge of the total picture, clear direction and change requirements. Thinking must be supported by a strong organizational will or intention, persuasive and convincing in nature. A strong motivational factor attitudinal, and necessary pre-requisite for success [10]. Action refer to action manage by continuous control and monitoring and review. Key performance indicator (KPI) is a necessary tool to measure stability and success of the strategic planning implementation.

Planning is a process, and process has to be able supported by documents and measurement for effective implementation. Process in planning include setting of goals, determining action, allocation of resources and building measurable performance indicators and a positive outcome may be realized in the form of strong rapport, cooperation and collaboration[17]. There is a need for better management control and planning function for managers[17][18]. They identify 29 key issues between 1996 and 2000 and among top ten issues is management focus. Three top five of the issues are related to ICT strategies as follow:

- a. Improving information security including data integrity and quality
- b. Managing integration of data processing, office automation and telecommunication
- c. Aligning IT with the enterprise
- d. Using IT competitive advantage
- e. Improving IT strategic Planning

4.5 Malaysian Public Sector ICT Strategic Plan

In the 8th Malaysian plan, information Technology Strategic Plan (ISP) from Agencies is essential as planning document for Government Internet & IT Secretariat to develop Information Technology policies in public sector in line with the National IT Agenda. Malaysian Administration & Man Power Planning Unit (MAMPU) has develops a framework for to assist Agencies to develop their ISP. Currently, Malaysian Public Sector ICT Strategic Plan is a blueprint developed for agencies for the use of ICT, which is both cost effective and aligned with prioritized management and operational needs. This plan will lead to the identification of strategic initiatives, either new or already underway, which can be aided by application systems, technology and governance. It contains several elements that drive the development of the plan; business alignment, vision, allocation and value, enabling, governing and capabilities, and finally action and measurement. An ICT Strategic Plan will address the alignment between the strategic operational requirements of an agency enterprise and the planned and economically justified application of technology assets. This alignment is best achieved when ICT has been deployed as an enabler of one or more cross-functional business processes. The basic interplay between the business strategy area of expertise and the ICT strategy area of expertise is the dynamic between business "requirements push" and "technology pull". The "requirement push" is the direct specification of information and processing needs by the business. However, true competitive advantages are very often gained through "technology pull". This occurs when the business is made aware of opportunities through insight and understanding of the capabilities of technology. Without this insight there may be no customer self-service, costly information exchange, and a disadvantage in establishing eBusiness and eCommerce[19].

The ICT planning approach for Malaysian Public Sector is shown in figure 5.0 below. The methodology adopts the 4-stage approach that answer the why, what, how and when questions for each activities of an ICT strategic Planning formulation. The stages are subdivide into phases as figure 5.0 and each phases contains steps and detail activities. ICT strategic plan outputs are produce in every phase of this methodology, and categorized into deliverable and interim outputs. Deliverables are provided to the agency for approval, usually making completion of stage and are formal project document. On the other hand, interim outputs are internal project team document that support the mark progress toward the deliverable.

5.0 Information Security Defined

Information security defined as an information security program that plan to mitigate risk associated with the

processing of information in three elements which defined by security professional [2]:

- **Confidentiality.** The prevention of unauthorized use or disclose of information. Privacy is closely related topic that has lately been getting more and more visibility
- **Integrity.** Ensuring that information is accurate, complete, and has not been modified by unauthorized user or process.
- **Availability.** Ensuring that the users have timely and reliable access to their information assets.

These three elements (CIA) are the basics around which all security programs are developed. Both of them are linked together in the idea of information protection. The main idea is to show that information is an asset that requires protection.

6.0 Public Sector ICT Security

Public Sector ICT Security can be defined as the process of ensuring business continuity and services provision free from unacceptable risk. It also seeks to minimize disruptions or damage by preventing and minimizing security incidents[20]. The security of information within the Government of Malaysia's Information and Communications Technology (ICT) system is a subject of major concern. Threats such as impersonation, malicious code, misuse of data, easily available penetration tools, and powerful analytical techniques contribute in whole or in part to the necessity of providing adequate protection to public sector ICT assets. These threats if left unchecked, will result in painful explanation at the very minimum or untold damage to the country.

Over the years, government agencies have been religiously collecting vast amount of information. It is in the early 70's that these information have been deposited into digital format and since then, these repositories have unknowingly become exposed because of the invaluable information they keep and now in a format easily manipulated without stringent audit trail. The government realizes this and that the government is also aware that there is an urgent need to secure the vast information resource through effective management of the security of ICT systems. In this regard, efforts are being made to ensure Public Sector ICT Security management achieve and maintain a high level of confidentiality, integrity and availability. A comprehensive approach is required in planning, developing, operating and maintaining the government's ICT security processes. The ICT security measures need to be incorporated early, in the requirement specification and design of the ICT system, before the implementation stage to ensure a cost-effective and comprehensive system.

The main steps are assessing the current security strengths and vulnerabilities; developing ICT security policies, standards and processes; designing and develop customized security architecture; and evaluating and selecting the best security system for the organization.

The ICT security process should mirror the management's direction in relation to overall organizational policy; organizational roles and responsibilities; personnel; government's asset classification and control; physical security; system access controls; network and computer management; application development and maintenance; business continuity; compliance to standards as well as legal and statutory requirements; classification and protection of information media; employee awareness programs; and incident reporting and response[20]

7.0 Integrating the Strategic IS/IT Planning with Information Security

Integrating both Strategic IS/IT planning process with the Information Security Domain will produce more valuable SISP to the organization. The appropriate domain should include in the SISP process as an input in the external business environment, external IS/IT environment, internal business environment and internal business environment. As for example, security management practice domain should include in all input attributes; security management in internal and external business and IS/IT environment. Beside SISP and Information Security Domain, Computer System Life Cycle should be incorporate in the establishment of SISP, especially in the development or acquisition phase. At this phase, security requirement was identified then incorporate security requirement into specification. Without policy, security practice will be developed without clear demarcation of objectives and responsibilities[21].

8.0 Conclusion

Base on the discussion above, there is a need to ensure Public Sector ICT security management achieve and maintain a very high level of confidentiality, integrity and availability in their information systems which can be plan at the early stage. The most appropriate time to plan the security when the information systems plan is conducted. Finally, the strategic security planning should align with the strategic information systems plan.

References

- sector". PhD. Dissertation, Universiti Putra Malaysia, 2004
- [1] Doherty, N. F., Marples, C. G. & Suhaimi A. "The relative success of alternative approaches to strategic information systems planning: an empirical analysis". *Journal of Strategic Information Systems*. Vol 8 pp 263-283.
- [2] Earl, M. J. "Experience in strategic information systems planning". *MIS Quarterly* Mac, 1993, pp1-24.
- [3] Hevner, A. R., Berndt D. J. & Studnicki J. "Strategic information systems planning with box structures". *IEEE Proceeding of the 33rd Hawaii International Conference on Systems Science*. Hawaii, January 2000.
- [4] Isa, M. "Strategic Information Systems Planning in Islamic Information Systems: A case study for Malaysian Muslim Organization". Ph. D. dissertation, University of Wales, Lampeter. 1997
- [5] King, W. "Creating A Strategic Capabilities Architecture, Information Systems Management", 1995, v.12.
- [6] Wylder, J. "Strategic Information Security", Auerbach Publication, 2004.
- [7] Ward, J. & Griffith, P. "Strategic Planning For Information Systems" (2nd Edition). John Wiley & Son, London. 1996.
- [8] Robson, W. "Strategic Management & Information Systems" (2nd Edition), 1996, Pitman Publishing, London
- [9] Lederer, A., L., Sethi, V. "Seven Guidelines For Strategic Information Systems Planning, Information Strategy: The" *Executive's Journal*, 07438613, Fall98, Vol. 15, Issue 1, p23, 6p, Auerbach Publications Inc, 1998.
- [10] Malik, R. Preparation of Information technology strategic planning. MAMPU-Intan IT Manager Seminar, INTAN, Kuala Lumpur, 2001. Pp1-11.
- [11] Remanyi, D. Strategic Information Systems, Current Practice and Guidelines, Ph. D. dissertation, Henley Management Collage, 1991, UK.
- [12] Muhamad Nubli. "Source-agent-client component in strategic information systems planning methodology for Malaysian public
- [13] Zurina, U., Hassan, M., Mamat, A. & Izura, N. "Kajian Pelaksanaan Perancangan Strategik Sistem Maklumat di Organisasi", *Proceeding on REDEC 2001 colloquium*, UNITEN, Kajang pp 24.
- [14] Maslinda, N. Hassan, M. & Mamat, A. "Perancangan strategic system maklumat: factor mempengaruhi kepuasan perancang. *Proceeding on REDEC 2001 colloquium*, UNITEN, Kajang pp 15
- [15] Selamat, M. H, Alias, R. A., Ibrahim, O. & Abdullah, S. "IT Strategic Planning For School Resource Center. *Proceeding of the 13th. Educational Technology Convention*, 19-21 September, 2000, Hotel HillCity, Ipoh, Perak, Malaysia.
- [16] Alinda, R. A., Selamat, M. H., Abdullah, S. & Ishak, I. S. "Strategic Information Systems Planning for IHLs: A preview", *Proceeding of the Malaysian Science and Technology Congress 2001 (MSTC 2001), Information and Communication Technology Session*, Universiti Science Malaysia, 8-10 Nov, 2001, Malaysia.
- [17] Suraya, M. "Developing a prototype for starategic information systems planning toolkit". 2001, M. Sc. (Computer Science) dissertation. UTM
- [18] Md Zabid Abd Rashid, Han Chun Kwong, & Ronald M. Zigli. "Key Issues in MIS in Malaysia", 16-17 Oct, 1995.
- [19] MAMPU. "Standard, Policies and Guidelines – Malaysian Public Sector ICT Strategic Planning Guide", Version 1.0, Putra Jaya. August 2003.
- [20] MAMPU. "Malaysian public sector management of information & communication technology security handbook (MyMIS)". Putra Jaya., 2002
- [21] Higgins, N. H. "Corporate Systems Security: Towards an integrated management approach". *Information Management & Computer Security*, 1999; 7(5) pp217-222

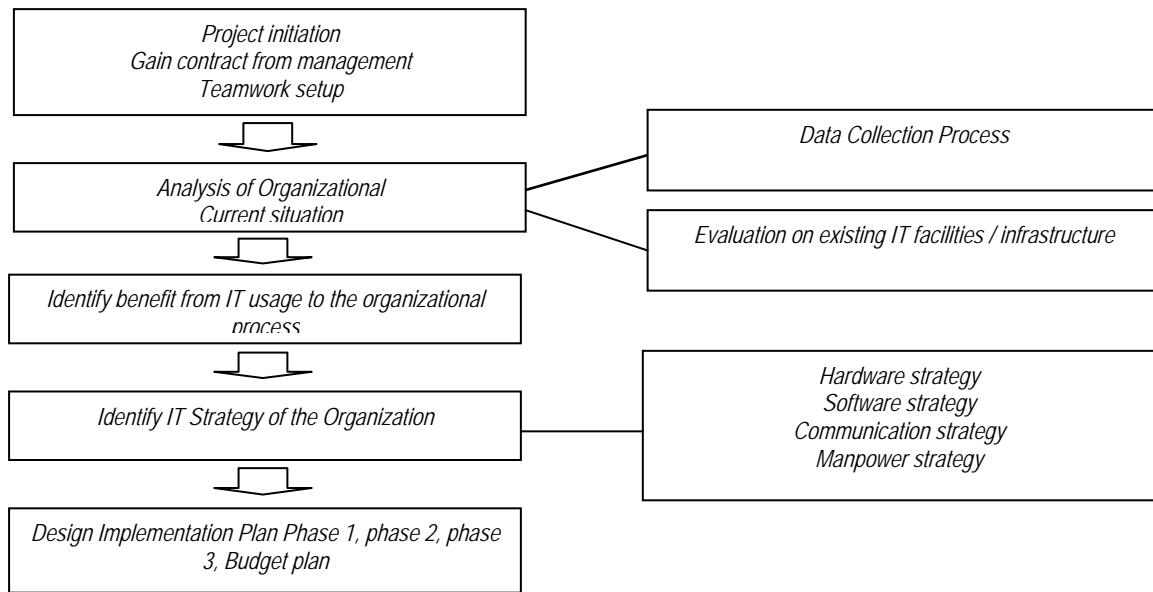


Figure 2.0: Process Involved in SISP Methodologies (Muhammad Nubli, 2004)

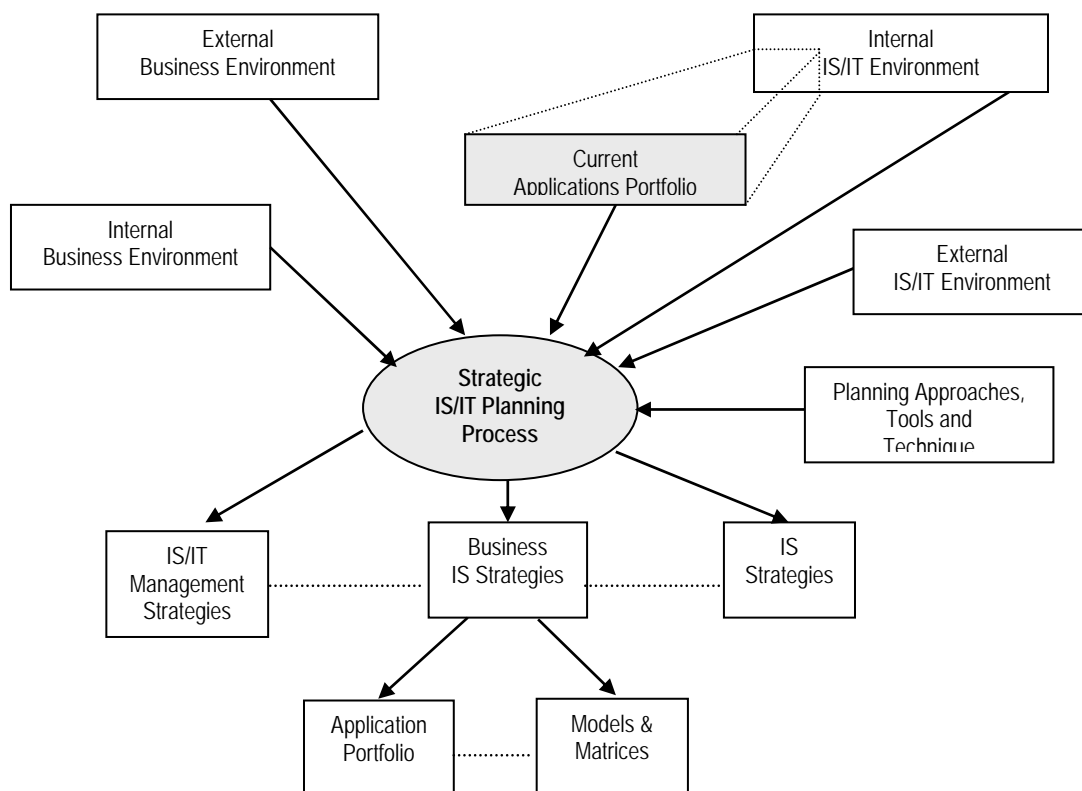


Figure 3.0 The inputs and outputs of the planning process (Ward, 1996)

A Preliminary Study on Embedded Platforms for Computer Vision Applications

Kenny Kuan Yew KHOO, Yong Haur TAY

Universiti Tunku Abdul Rahman
13, Jalan 13/6. 46200, Petaling Jaya, Selangor
kennyhooky@gmail.com

ABSTRACT

Today, silicon chips have become more and more powerful despite reduction in size. Various general processors, digital signal processors (DSPs) and graphic processing units (GPUs) have been embedded in countless electronic gadgets such as Personal Digital Assistants (PDAs), mobile hand-phone, digital cameras etc. With the high processing power of these embedded platforms and ever increasing size of both volatile and non-volatile storage, it opens up a great opportunity to integrate computer vision (CV) applications to these relatively low-cost and stand-alone devices.

Applying CV applications inside electronic gadgets can be the next trends in industrial or commercial market. These devices can do specific smart and intelligent job, such as object or text recognition for manufacturing, surveillance and security, as well as entertainment.

We found that there is still yet to have a standard embedded software or hardware architecture in embedded systems (ES) for CV. An open architecture is needed for CV researchers to develop CV application without worrying the underlying embedded platform or CV library dependency in Embedded System (ES).

This paper provides a preliminary study of embedded computer vision (ECV) and tries to develop a possible framework for ECV. We also suggest a standard/open embedded platform for all the CV applications, which share the same embedded system characteristics and CV library.

KEYWORDS

Embedded systems, computer vision, embedded computer vision.

1. Introduction

In the last decade, we have witnessed a dramatic increase in the use of CV in ES. Recently CV was successfully implementing in various mission-critical systems, such the landing of the mars rovers to computer-aided surgery [1].

Different computer vision domains such as text recognizer and surveillance were addressed separately due to the different behavior of the application. However, these domains share many common problems related to their embedded-system characteristics such as real-time, memory and semaphore. CV researchers have difficulties to implement their CV algorithm into a small embedded device for a specific CV application. These situations occur because there are:

- (i) No standardized framework or architecture in embedded system for CV and
- (ii) No common CV programming library

CV is also widely used in industrial process and control. It is mainly use for product inspection to replace human eye, for exhaust and long hour inspection processes. Found that cameras module is in everyday appliances such as cell phones and PDA. The appliances are becoming “smarter” by gaining the capability of processing the acquired images internally.

The CV field can be distinguishing as immature and diverse. There is no standard procedure on how computer vision problems can be solved. Instead, there is abundance of methods for solving various well-defined computer vision problems.

Today mostly ES are the system on a chip (SOC). Which the CPU was purchased as intellectual property to add on the IC's layout as an application-specific integrated circuit. For example:-

- (i) Integrate a video processing unit on chip for multimedia purposes, such as Sigma design

Inc.'s multimedia chip EM8600 [2] which can decode High Definition Windows Media Format 9. The equivalence processing power in this video encoding is only found in Intel Pentium 4 3.06GHz. [3]

- (ii) Integrate voice codec with ARM processor for VOIP purpose such OCTASIC which can handle 672 channel of G.971 audio codec at a time [4].

Notice that the intellectual property in the SOC is actually a unit to perform a specific task in real time where general purpose microcontroller can't perform. Currently there is no specific SOC for CV application, because the CV methods often are very task specific and seldom can be generalized over a wide range of applications. Mostly ES for computer vision is laid on platform of combination between microprocessor and DSP chip. The DSP chip can perform special calculation which can be 1,000 to 10,000 times much faster than microprocessor.

Combining CV and ES eventually create a new field, called "embedded computer vision" or ECV. A growing demand for ECV applications exists to tackle problems which can hardly be solved by traditional sensor systems, i.e. camera. However, due to the high computational complexity of most computer vision algorithms, realization of such applications on embedded systems is not a straight forward task.

This new field becomes important since more people show interest in ECV, and attempt to bring more CV applications in ES form. IEEE organization also shows interest by conducted the first embedded computer vision workshop in Jun 2005 at San Diego, USA. Hence this ECV survey tries to study and develop a possible framework for ECV. It also suggest an open architecture of embedded platform for all the CV applications, which share embedded system characteristics and CV library.

Particularly, this survey will address the following questions:

- (i) What are the current research problems and the applications within computer vision domain, which are specific to embedded systems (for example, algorithms for efficient utilization of embedded hardware for computer vision)?
- (ii) What are the specific issues in embedded systems domain, which are relevant to computer vision (for example, meeting real-time constraints in embedded computer vision systems)?

2. Computer Vision Domain

The CV field can be characterized as immature and diverse, and there are many formulations to solve a particular CV problem. CV methods often are very task specific and seldom can be generalized over a wide range of applications. Typical computer vision system can be generalized into the following subsystems:

- (i) Image acquisition
- (ii) Preprocessing
- (iii) Feature extraction
- (iv) Recognition

The image acquisition subsystems are just a simple communication protocol to grab image or image sequence from imaging devices, e.g. camera, radar, tomography system.

Preprocessing subsystem use a lot of memory to do image processing for noise reduction on the image, it also reduces the overall amount of data.

Feature extraction subsystems perform a lot of CV algorithms for further reduce the raw data to a set of features, which ought to be invariant to disturbances such as lighting conditions, camera position, noise and distortion. Examples of feature extraction subsystems are:

- (i) Performing edge detection or estimation of local orientation.
- (ii) Extracting corner features.
- (iii) Detecting blob features.
- (iv) Extracting spin images from depth maps.
- (v) Acquiring contour lines and curvature zero crossings.
- (vi) Generating features with the Scale-invariant feature transform.

After getting a set of features, CV researchers are able to apply own idea/algorithm for specific CV application. The recognition step has to bring up a final hypothesis. Below are some examples of registration methods:

- (i) Least squares estimation
- (ii) Hough transform in many variations
- (iii) Geometric hashing
- (iv) Particle filtering
- (v) RANdom SAMple Consensus

Typical tasks of CV systems application that currently realize, can be categorized as:

- (i) Object recognition - detecting the presence of known objects or living beings in an

- image, possibly together with estimating the pose of these objects.
- (ii) Optical character recognition (OCR) -takes pictures of printed or handwritten text converts it into computer readable text such as ASCII or Unicode.
 - (iii) Tracking - known objects through an image sequence
 - (iv) Scene interpretation - create a model from an image/video of scenes.
 - (v) Ego motion - determine the motion of the camera itself.

The CV subsystems, typical task of CV and novel CV functions can be a referent point to generalize all the CV method/function over a wide range of applications to be a standard CV library.

Based on each CV subsystems, there are using different amount of computing systems resource, as shown in the following table. Some CV applications always run in critical time, any data which can not process in real-time will be ignored. Bottleneck of any CV subsystems might cause real-time processing fail. Preprocessing subsystems always deal with adjustment of large amount images. And feature extraction subsystems processing are worst, because it extracts a set of attribute from each static image or a sequence of images. Both subsystems are computation intensive and are dealing with large amount of data. Only DSP/GPU which have SIMD capability can parallel process all the data in real-time.

CV subsystems	Computing systems resources			
	I/O	DSP	Memory	Processor
Image acquisition	Yes		low	Low
Preprocessing		Yes	very high	High
Feature extraction		Yes	very high	very high
Registration	Yes		high	High

Table 1: Computing systems resources usage according to computer vision subsystem, in terms of I/O, DSP, memory and processing time

2.1 Embedded Systems Domain

In the beginning of embedded systems evolution (Figure 2), ES is just an input output device which does give electrical output base on incoming command, or a variable, the software architecture just a control loop systems.

At the time when microprocessor become fast and arithmetic logic unit (ALU) become more advanced, ES itself can collect data, and perform

calculation to determine the decision and output. At this stage, ES can contain a simple operating system and software structure.

When sound and image processing are needed inside ES, it involves large data manipulation. For example, a simple brightness adjustment operation for an image needs to perform 1 million time color.brightness function for each pixel in 1280 x 1024 pixel images.

$$\sum_{X=0}^{1280} \sum_{Y=0}^{1024} color.brightness(RGB, 10\%)$$

Figure 1: Example for calling color.brightness function to reduce 10% brightness of the image

When involve a sequence of image and sound, it is called video processing. Due to the large amount of data to be processed, microprocessor itself might not be able to processing this video data in real time. Integrated DPS/GPU in the systems helps to solve this video data processing in real time. DSP/GPU is able to processing in parallel.

The sequence of images and sound is called video. Microprocessor it self is might not be able to process this video data in real time. Integrated DPS/GPU in the systems helps to process this video data in real time. DSP/GPU is able to do parallel processing, and a single instruction for multiple data operations to solve the large data calculation

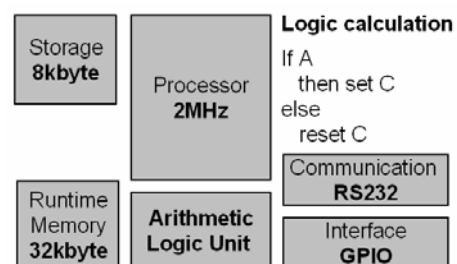


Figure 2: Early state of ES evolution

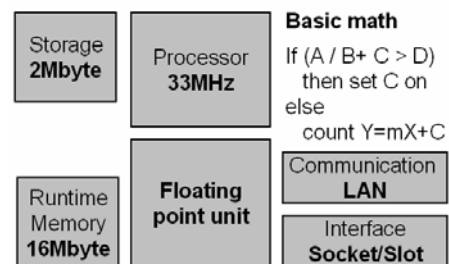


Figure 3: Middle state of ES evolution

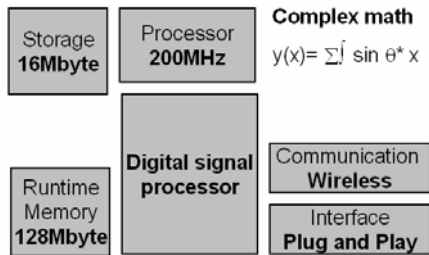


Figure 4: Current state of ES evolution

In embedded systems evolution shows us that when the time goes on, embedded systems are capable to contain an OS and cooperate with DSP/GPU to calculate complex math. With the high processing power of these embedded platforms and ever increasing size of both volatile and non-volatile storage, it opens up a great opportunity to integrate computer vision (CV) applications to these relatively low-cost and stand-alone devices.

Today LINUX and Windows CE battle on embedded system software, because embedded hardware growth will be at the aggregate rate of 14.2% to reach \$78.7 billion in 2009, while embedded board revenues will increase by an aggregate rate of 10% [5]. There are many different CPU architectures used in embedded designs [6], and the most popular embedded processors architectures are ARM, MIPS, PowerPC, and x86.

For very-high-volume embedded systems, a common systems configuration is the system on a chip. It is also an application-specific integrated circuit, where the CPU/DSP was purchased as add-on intellectual property inside the IC's design. A related common scheme is to use a field-programmable gate array, and program it with all the logic, including the CPU. Most modern FPGAs are designed for this purpose.

To support CV application, ECV platform must have at least an image sensor device and complex calculation capability.

3.1 Image Sensor Devices

Constraints in image sensor are sensitivity, resolution, progressive/interlaced scan image and interval between sequence images (video frame per second). These constraints will determine the size of information that CV systems receive for analysis. An ECV hardware platform must select the right image sensor, because it will provide sufficient information for CV systems to calculate accurate result.

Image sensor device provideS static or sequence image data. Such:

- (i) Camera (color image)
- (ii) Radar (Radio Detection and Ranging)
- (iii) LIDAR (Laser Imaging Detection And Ranging)
- (iv) Tomography system (imaging by sections or sectioning)

Camera is the most popular image sensor in CV application. Nowadays camera is getting a part in everyday appliances such as cell phones, PDA's, presentation appliances and vehicles. Moreover, cameras are becoming "smarter" by gaining the capability of processing the acquired images internally.

Radar and LIDAR image systems use radio/laser waves to determine the distance to an object by measuring the time delay between transmission pulse and detection of the reflected pulse. Both are rarely use in CV application, because both systems is normally target for far object which is invisible on eye.

Tomography involves the generation of a two-dimensional image representing a slice or section through a three-dimensional object.

3.2 DSPs and GPUs

A digital signal processor (DSPs) is a specialized microprocessor designed specifically for digital signal processing, generally in real-time. DSP functionality can also be realized using Field Programmable Gate Array chips.

Most DSPs use fixed-point arithmetic, because in real world signal processing, the additional range provided by floating point is not needed, and there is a large speed benefit; however, floating point DSPs are common for scientific and other applications where additional range or precision may be required.

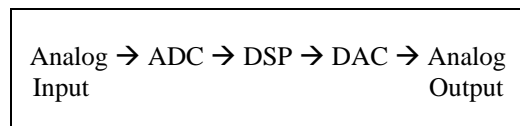


Figure 5: Processing of digital data

If using DSP in ECV platform, ADC or DAC component of DSP will be ignored, because CV application mostly import image from image sensor module (camera module). DSP is suitable to perform variety of CV algorithms.

GPUs differ in some major ways from DSPs architectures [7]. All computation is performed

with floating-point arithmetic; currently, there are no bit or integer math instructions. Also, since the GPU is designed to work with images, the memory system is in effect a 2D segmented memory space—a segment number (i.e. an image to read from) and 2D address (the x,y position in the image).

Moreover, there are no indirect write instructions. The output write address is fixed by the rasterizer and cannot be changed by our program. This can be particularly challenging for algorithms that naturally scatter into memory. Finally, no communication is allowed between the processing of the different fragments. In effect, the fragment processor is an SIMD data-parallel execution unit, independently executing our code on all fragments.

Despite these constraints, a variety of CV algorithms can be efficiently implemented on the GPUs, ranging from linear algebra and signal processing to numerical simulation. Though conceptually simple, computing with GPUs can be frustrating for first-time users because of the need for graphics-specific knowledge. But some software tools can help. Two high-level shading languages, CG and HLSL, let users write C-like code that compiles to fragment program assembly. Though the languages greatly simplify the writing of shader assembly code, applications still must use the graphics API to set up and issue computation.

4. Current Related Activity

Currently there are two main workshops which aimed to bring together researchers working on computer vision problems, which share embedded system characteristics. This workshop also is an area to exchange ideas and experiences.

- (i) Workshop on Intelligent Solutions in Embedded Systems (WISES)
<http://www.vmars.tuwien.ac.at/wises>
 From: Vienna University of Technology, AUSTRIA, Since 2003
- (ii) Embedded Computer Vision Workshop (ECVW)
<http://www.scr.siemens.com/ecv05/>
 From: IEEE, Since; 2005

A lot CV related project from academic and open source community. For this ECV systems survey, we are look closely to a CV related project which direct implement into hardware and least dependent to any library or systems. A lot of CV related project is base on JAVA and C# language, it is target for desktop PC processing, and more toward CV algorithm development.

- (i) OpenVIDIA: Parallel GPU Computer Vision
<http://opencv.sourceforge.net>
 From university of Toronto
 Latest release version 0.8.3 (Mar 23, 06)
- (ii) OpenCV: Open Computer Vision Library
<http://www.intel.com/technology/computing/opencv/index.htm>
 From: INTEL
 Latest release: version 2.3.0 (Aug 3, 05)
- (iii) VXL: Vision-X-Library
<http://vxl.sourceforge.net>
 From: Open Source Community
 Latest version 1.4.0 (Jan 11, 06)
- (iv) VIGRA: Vision with Generic Algorithms
kogs-www.informatik.uni-hamburg.de/~koethe/vigra/
 From: University of Hamburg, Germany
 Latest version 1.4.0 (Dec 21, 05)

During the past few years there has been a large and energetic upswing in research efforts aimed at synthesizing fuzzy logic with neural networks. Neural networks provide algorithms for learning and are modeled after the physical architecture of the brain. Fuzzy logic deals with issues such as reasoning at the semantic or linguistic level and is based on the way brain deals with inexact information.

A variety of fuzzy-neural network models have been used in computer vision. Consequently, the two technologies complement each other.

- (i) Lightweight Neural Network++
[\(http://lwneuralnetplus.sourceforge.net/\)](http://lwneuralnetplus.sourceforge.net/)
 From: Open Source Community
 Latest release:lwneuralnet++ 0.998 Dec 27, 2005
- (ii) NNF - Neural Net Framework
<http://sourceforge.net/projects/nnf>)
 From: Open Source Community (GNU Library or Lesser General Public License (LGPL))
 Latest release:CVS only BETA
- (iii) FANN - Fast Artificial Neural Network Library
<http://leenissen.dk/fann>)
 From: Open Source Community
 Latest release:fann-2.0.0.tar.bz2 (Jan, 7 2006)

Commercial CV products for industry, is no attempt to cover image processing or any of the many suppliers of sensors, frame grabbers, or other equipment to the industry. The total size of the worldwide computer vision industry is estimated at \$6.6 billion (2003).[8]. Most industry CV product focus on manufacturing, such as circuit board and semi-conductor wafers

- (i) KLA-Tencor <http://www.kla-tencor.com>
Revenues: \$1.3 billion (2003),
Vision technologies: Uses morphological processing among other vision methods to match circuit patterns to a database for defect detection.
- (ii) Orbotech <http://www.orbotech.com/>
Revenues: \$228 million (2003).
Vision technologies: Seems to mostly use 2-D feature detection and registration to compare images of printed circuit boards to a master image.
- (iii) Matrox <http://www.matrox.com/>
Revenues: \$500 million (2001)
Vision technologies: provide standard template matching and measurement functions.

Currently commercial CV products for consumer is least, it is just target for security, and entertainment. Security camera is added computer vision application, such event-controlled recording upon detecting movements, such MOBOTIX's M22M Camera [9]. Entertainment product such SONY AIBO is apply computer vision system inside this robot dog toy, so that this toy is able to walk and "see" its environment via camera [10].

5. Embedded Computer Vision Systems

Basic idea behind ECV systems is to implement a specific CV task inside a box of ES which is low cost, standalone, and portable.

ECV systems will allow CV application engineer to develop without worrying the underlying ES platform or CV library dependency. Based on current state of observe, constraint for develop ECV systems should contain:

- (i) Systems workflow
- (ii) Systems architecture
- (iii) Systems benchmark

This ECV systems workflow can be divided into 3 main parts:

- (i) CV application development
- (ii) ECV systems implementation
- (iii) ES design

In ECV systems, there is a need to generalize CV function/algorithm into a novel standard CV library. And CV developers must use the only the functions in ECV library to develop CV applications. Since the field of CV is immature and diverse, there is possible of any new method/algorithm is discover later. Only CV researcher who's also know well in

ES/DSP/GPU field can implement or integrated new CV function in ECV library.

Any CV application base on ECV systems should be able to be benchmarked. And based on the benchmark index, ES designer should be easy to provide equivalent ES platform to meet the specific CV application.

To achieve the suggested ECV systems workflow, this survey will:

- (i) Study the possible to generalize CV algorithm into a standard CV library.
- (ii) Study the scope of a ECV implement engineer
- (iii) Study an benchmarking issue in ES cross-platform and CV application

These ECV systems must meet their real-time constraints with a combination of special purpose hardware, DSP/GPU chip and software tailored to the system requirements. Suggest ECV systems architecture such as figure 6, show that ECV library should be directly optimize with DSP/GPU chip.

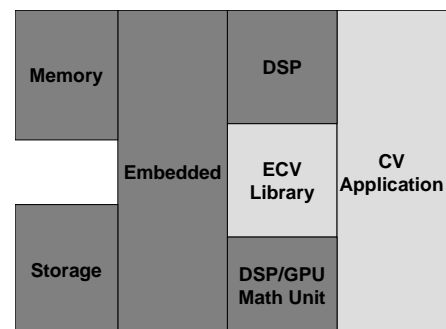


Figure 6: ECV systems architecture

According to table 1, ECV systems should contain below component inside each subsystems:

- (i) Imaging system driver for Image acquisition subsystem
- (ii) Imaging processing library for preprocessing subsystem
- (iii) Standard CV function set for Feature extraction subsystem
- (iv) Collection of CV application set for registration subsystem

In computing, a benchmark is the result of running a computer program, or a set of programs, in order to assess the relative performance of an object, by running a number of standard tests and trials against it. This survey might study implement synthetic type of benchmark into ECV systems to take statistics of all type of CV function/operations from plenty of application programs in various ECV platform

and get proportion of each operation in each ECV platform. The benchmark results will represent in KWIPS (kilo whetstone instructions per second). And this value will be useful to determine the optimize ECV hardware specification for the specific CV application

6 Conclusions

Currently there are some CV libraries for PC environments, but they do not directly support embedded systems.

While this project will identify the trends in DSP/GPU and their computational models, the project will also study a best structure of CV library to implement in their DSP/GPU model

As a result, the project might propose a framework or architecture for ECV software and hardware so that both of them cooperate and utilizes the computational power of CPU and DSP/GPU unit to perform real-time CV. At the same time leaves the CPU free to conduct other tasks beyond vision.

7. References

- [1] Larry Matthies, Jet Propulsion Laboratory, "Vision Systems for Mars Rovers and Landers", <http://www.scr.siemens.com/ecv05>
- [2] Sigma Designs Press Release, "I-O Data Launches Networked DVD Media Player With WMV9 Support", http://www.sigmadesigns.com/news/press_releases/040824.htm
- [3] Microsoft, WMV HD Content Showcase, "High Definition Quality with WMV9", <http://www.microsoft.com/windows/windowmedia/musicandvideo/hdvideo/contentshowcase.aspx#sysreq>
- [4] Robert Keenan, CommsDesign.com, "Startup Octasic delivers scalable echo canceller", <http://www.commsdesign.com/news/OEG20030407S0070>
- [5] Ravi Krishnan, BCC, "Future of Embedded Systems Technology", http://www.electronics.ca/reports/embedded/systems_technology.html
- [6] LinuxDevices.com, "Embedded Processor and System-on-Chip Quick Reference Guide", <http://www.linuxdevices.com/articles/AT4313418436.html>
- [7] Ian Buck, Researcher, Graphics Laboratory, Stanford University, "Programming GPU for general computing", http://www.eetasia.com/ART_8800367818_480100_eaad0254200506_no.HTM
- [8] David Lowe, University of British Columbia, "The Computer Vision Industry", <http://www.cs.ubc.ca/spider/lowe/vision.htm>
- [9] MOBOTIX AG, "The new face of IP video", <http://www.mobotix.com/content/view/full/2>
- [10] Wikipedia, the free encyclopedia, "AIBO", <http://en.wikipedia.org/wiki/AIBO>

Modeling Wind Speed Data in Surabaya and Sumenep, Indonesia Using Vector Autoregression (VAR)

Irhamah^{1*}, Eling Anindita², Dwi Endah Kusri³

Department of Statistics, Institute of Technology Sepuluh Nopember,
Kampus ITS Keputih, Sukolilo Surabaya, 60111, East Java, Indonesia
e-mail: irhamahn@yahoo.com

¹PhD student at Department of Mathematics, Faculty of Science, UTM, Malaysia

Abstract

As one of important meteorology parameters, wind speed is very useful to detect air pollution, for the need of take off and landing of airline, weather forecast, metallurgy, irrigating, study of land friction, mining, manufacturing, industrial, and early warning system for natural disaster. Therefore the study of wind speed phenomenon is needed, especially to know its forecasting model.

Research about modeling wind speed in two bunch up town in Indonesia that is Surabaya and Sumenep, have been conducted but still in univariate time series way. In this paper, the use of multivariate time series method that is Vector Auto Regression (VAR) for modeling wind speed at Juanda - Surabaya and Sumenep – Madura was presented, and also for finding relationship between these two variables. The data used here was the daily mean of wind speed data (measured in knots) for 6 months (September 2004 until February 2005).

The first step in data analysis was identification that started with the examination of stationarity assumption based on descriptive statistics, Autocorrelation Function (ACF) plot and Matrix ACF plot. Examination results indicated that there were non-stationarity in mean and variance, so that first differencing was done to overcome non-stationarity in mean and root-square transformation was done to overcome non-stationarity in variance. Afterwards VAR method was applied to the stationer data.

Modeling with VAR to the wind speed data in Surabaya and Sumenep yield the best model VAR (4). In the form of equation, the model can be written as following.

$$Sby_t = [0.663\sqrt{Sby_{t-1}} + 0.118\sqrt{Sby_{t-2}} + 0.219\sqrt{Sby_{t-3}}]^2$$

$$Sum_t = [0.611\sqrt{Sum_{t-1}} + 0.218\sqrt{Sum_{t-2}} \quad 0.041\sqrt{Sum_{t-3}} + 0.059\sqrt{Sum_{t-4}} +$$

$$0.170\sqrt{Sby_{t-4}} \quad 0.170\sqrt{Sby_{t-5}} + 0.153\sqrt{Sum_{t-5}}]^2$$

Beside affected by itself at one, two, three, four, and five days before; wind speed in Sumenep also affected by wind speed in Surabaya at four and five days before, while wind speed in Surabaya is only affected by itself at one, two and three days before. The strong relationship between these two variables was also shown by a high measure of correlation between them.

Keywords: wind speed, time series, vector auto regression

1. Introduction

As one of important meteorology parameters, wind speed is very useful to detect air pollution, for the need of take off and landing of airline, weather forecast, metallurgy, irrigating, study of land friction, mining, manufacturing, industrial, and early warning system for natural disaster. Therefore the study of wind speed phenomenon is needed, especially to know its forecasting model.

Analysis and modeling of sequentially observations like wind speed, usually use time series analysis. Researches about modeling wind speed in Indonesia using time series analysis have been conducted by Nuryana [1] in Sumenep's wind speed and by Hartanti [2] in Surabaya's wind speed, each by implementing univariate time series method, that is ARFIMA (Sumenep) and ARIMA (Surabaya) respectively. These researches show relationship among wind speed at one period and previous periods, but not the relationship

between those two town's wind speeds. Ewing et al. [3] have examined the interdependence in time series wind speed data measured in the same location at four different heights using a multiple-equation system known as a vector autoregression (VAR).

In this paper we present multivariate time series method for modeling wind speed in Surabaya and Sumenep simultaneously using VAR method to know the relationship between wind speed in Surabaya and Sumenep, in a closed-loop system where enable the existence of interrelationship between 'input' and 'output' variables (Chatfield [4]). VAR also can obtain best model of wind speed pattern for the need of forecasting.

2. Vector Autoregression (VAR) Model

VAR model is implemented when each data variable doesn't affected only by itself at different time lag, but also affected by other variables in a close-loop system. If data stay in a closed-loop system, then we do not need to think about term of 'input' and 'output'. General model of VAR (p) can be written as follows: (Hamilton [5])

$$\begin{aligned}
 \mathbf{x}_t &= \Phi_1 \mathbf{x}_{t-1} + \Phi_2 \mathbf{x}_{t-2} + \dots + \boldsymbol{\varepsilon}_t \\
 (1) \qquad &= \sum_{j=1}^p \Phi_j \mathbf{x}_{t-j} + \boldsymbol{\varepsilon}_t
 \end{aligned}$$

2.1. Model Identification

In principle, identification in vector time series model is similar with identification in univariate time series. Matrix autocorrelation function plot and matrix partial autocorrelation plot are commonly used-tool for identification of vector time series Z_1, Z_2, \dots, Z_n . (Wei [6])

Time series plot, MACF plot and MPACF can indicate non stationarity in data. If time series plot does not show mean's change in time or in other words there does not seem to be a significant trend or any seasonal pattern in data, the data is said to be stationary in mean. For non stationary data, differencing is needed.

If the variance changes in time, the data is said to be non stationary in variance and therefore we re-express the data with transformation in order to stabilize variance. Box-Cox transformation is commonly implemented by calculating λ value.

2.2. Model Selection

In time series analysis, it is possible to possess more than one suitable models. AIC (*Akaike's Information Criterion*) is one of criterion that also considers the number of parameter involved in model. The best model should not only produce sum of residual near to zero but also be parsimonious (having few parameters). Following literatures from Jiang and Kitagawa [7] and Basri [8], the best model selection in this paper is decided by minimum AIC (Enders [9]).

2.3. Parameter Estimation

Least square method is used for estimating the parameters in this analysis. After that, significant parameters are selected by backward elimination method. Parameter estimation in multivariate time series could be done thru parameter estimation of univariate regression. (Kapetanios [10]). The general model of regression in the form of matrix:

$$\begin{matrix}
 y_1 \\
 y_2 \\
 \vdots \\
 y_n
 \end{matrix}
 =
 \begin{matrix}
 1 & x_{11} & x_{21} & \dots & x_{k1} \\
 1 & x_{12} & x_{22} & \dots & x_{k2} \\
 \vdots & \vdots & \vdots & \vdots & \vdots \\
 1 & x_{1n} & x_{2n} & \dots & x_{kn}
 \end{matrix}
 \begin{matrix}
 \beta_0 \\
 \beta_1 \\
 \vdots \\
 \beta_k
 \end{matrix}
 +
 \begin{matrix}
 \varepsilon_1 \\
 \varepsilon_2 \\
 \vdots \\
 \varepsilon_n
 \end{matrix}
 \quad (2)$$

in other word

$$y = X\beta + \varepsilon \quad (3)$$

During the parameter estimation phase, a minimization function is used to maximize the likelihood (probability) of the observed series, given the parameter values. In practice, this requires the calculation of the (conditional) sums of

squares (SS) of the residuals that is $\varepsilon' \varepsilon = \sum_{i=1}^n \varepsilon_i^2$

(Draper and Smith [11]).

2.4. The Backward Elimination

The Backward Elimination begins with full model including all variables, followed by evaluation of parameters whether there are insignificant parameters. If there are several insignificant parameters, the most insignificant parameter (parameter with largest p-value) will be dropped first. This step will be repeated iteratively until there is no insignificant parameter.

2.5. Diagnostic Checking

Diagnostic checking of time series model is usually carried out using examination of several assumption, which is multi normality assumption, white noise residual assumption and homogeneity of residual. If the model found could not fulfill one or more assumptions, another model should be established.

3. Data and Research Methodology

Data used in this research were secondary data from BMG Surabaya about daily mean of wind speed data in Surabaya and Sumenep for 6 months starting from September 2004 until February 2005. Methodology of analysis can be seen at Figure 1.

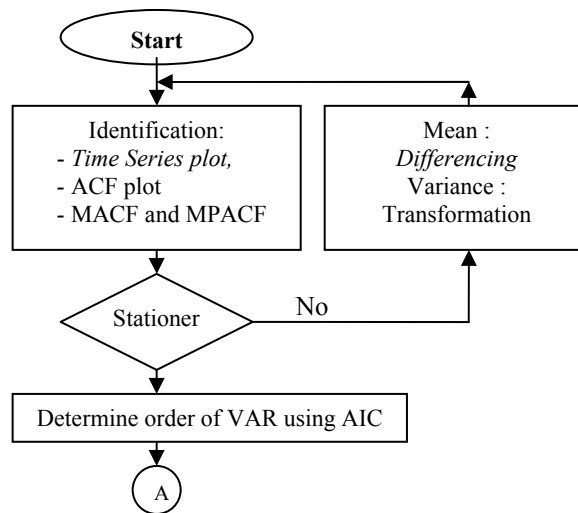


Figure 1. Flowchart of Research Methodology

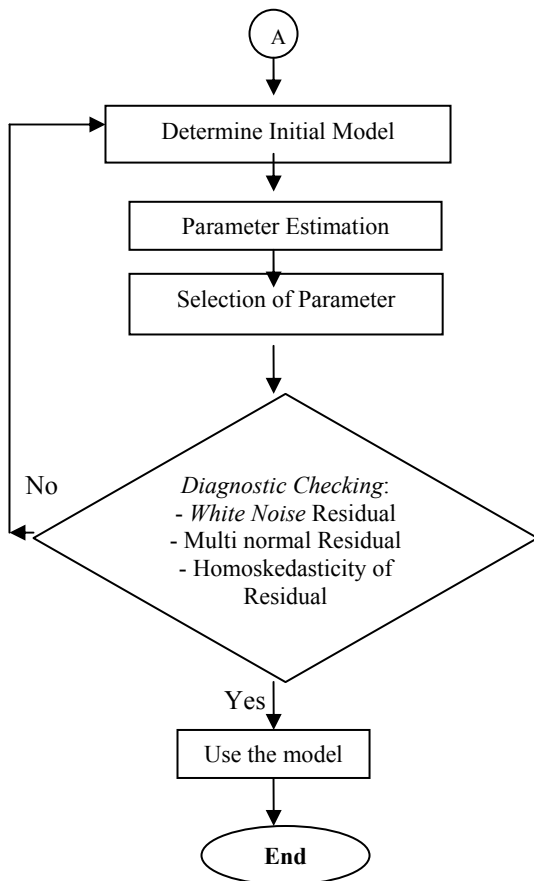


Figure 1. Flowchart of Research Methodology (continued)

4. Results and Discussion

4.1. Descriptive Statistics

Descriptive statistics are very useful to know patterns and characteristics of data, the results can be seen at Table 1.

Table 1. Descriptive statistics of daily wind speed data

	N	Min	Max	Mean	Std. Deviation
SURABAYA	181	2	13	6.51	2.11
SUMENEP	181	1	15	6.66	3.49

From Table 1, it is shown that the standard deviation of wind speed in Sumenep is higher than wind speed in Surabaya; it means that wind speed in Sumenep is relatively more fluctuate. This is could be caused by geographically position of Sumenep’s observation station which is located at the edge of beach. Wind speed at the sea or beach usually higher and more fluctuate than in the land, since the wind flow does not blocked by massif object like buildings or big trees. High fluctuation in Sumenep’s wind speed could also be caused by cyclone in south coastal area of Madura island (in particular at Camplong, Sampang and Kalianget, Sumenep) on January 2005 (Kompas, 11 January 2005). On the other hand, Surabaya’s wind speed is relatively constant since it is located near resident settlement and industry area.

4.2. Stationarity of Data

Time series plot of real data indicates that data contains non-stationarity in mean whereas the mean is varying through time. The non-stationarity in mean is also shown by its autocorrelation function that decays slowly in univariate ACF and multivariate MACF (Table 2), therefore differencing is done. Time series plot of real data can be seen at Figure 2.

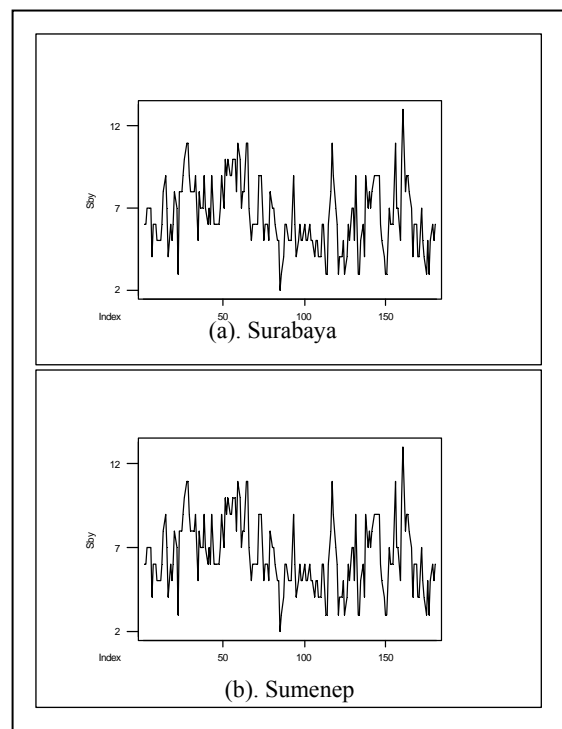


Figure 2. Time Series plots of wind speed in Surabaya and Sumenep

Table 2. Schematic representation of MACF

Name/Lag	0	1	2	3	4	5	6	7	8	9	10
sby	++	++	++	++	++	++	+	+	++	+	+
sum	++	++	++	++	++	++	++	+	+	+	+

+ is > 2*std error, - is < -2*std error, . is between

Table 3. Schematic representation of MPACF

Name/Lag	1	2	3	4	5	6	7	8	9	10
sby	+	+
sum	+	+

+ is > 2*std error, - is < -2*std error, . is between

It was also known from the time series plot that wind speed data in Sumenep was not stationer in variance. Therefore Box-Cox transformation was implemented to stabilize variance, optimal λ was found to be 0.5 (or square root transformation).

MACF and MPACF of root square-transformed and first-differenced data can be seen at Table 4 and Table 5.

Table 4. Schematic representation of MACF of transformed and differenced data

Name/Lag	0	1	2	3	4	5	6	7	8	9	10
sby	+	-	+
sum	..+	..--

+ is > 2*std error, - is < -2*std error, . is between

Table 5. Schematic representation of MPACF of transformed and differenced data

Name/Lag	1	2	3	4	5	6	7	8	9	10
sby	-	-	-
sum	..--	..+

+ is > 2*std error, - is < -2*std error, . is between

Schematic representation of MACF and MPACF of transformed and differenced data have shown that data already stationer in mean and variance.

4.3. Estimation of Initial VAR Model

Table 6. AIC Values

Lag	AIC	Lag	AIC
0	-641.971	6	-669.862
1	-667.367	7	-673.218
2	-670.452	8	-667.866
3	-672.075	9	-666.512
4	-674.562	10	-659.495
5	-673.915		

Minimum AIC was obtained by the 4th lag or VAR (4) that is -674.562, and from Table 5 it was shown that MPACF plot cuts off at the 4th lag. Hence the initial suggested model for wind speed in Surabaya and Sumenep was VAR (4).

4.4. Parameter Estimation of VAR Model

Objective of parameter testing phase is for finding suitable parameter involved in the model. Initial VAR model was VAR (4), which can be formulate as follows:

$$\begin{aligned} \sqrt{Sby_t^*} &= \varphi_{111} \varphi_{112} \varphi_{113} \sqrt{Sby_{t-1}^*} + \varphi_{121} \varphi_{122} \varphi_{123} \sqrt{Sby_{t-2}^*} + \\ &\sqrt{Sum_t^*} = \varphi_{211} \varphi_{212} \varphi_{213} \sqrt{Sum_{t-1}^*} + \varphi_{221} \varphi_{222} \varphi_{223} \sqrt{Sum_{t-2}^*} + \\ &\varphi_{131} \varphi_{132} \varphi_{133} \sqrt{Sby_{t-3}^*} + \varphi_{141} \varphi_{142} \varphi_{143} \sqrt{Sby_{t-4}^*} + \varepsilon_{1t} \\ &\varphi_{231} \varphi_{232} \varphi_{233} \sqrt{Sum_{t-3}^*} + \varphi_{241} \varphi_{242} \varphi_{243} \sqrt{Sum_{t-4}^*} + \varepsilon_{2t} \end{aligned}$$

Where

$$\begin{aligned} \sqrt{Sby_t^*} &= \sqrt{Sby_t} \quad \sqrt{Sby_{t-1}^*} = \sqrt{Sby_{t-1}} \quad \sqrt{Sum_{t-2}^*} = \sqrt{Sum_{t-2}} \quad \sqrt{Sum_{t-3}^*} = \sqrt{Sum_{t-3}} \\ \sqrt{Sum_t^*} &= \sqrt{Sum_t} \quad \sqrt{Sum_{t-1}^*} = \sqrt{Sum_{t-1}} \quad \sqrt{Sby_{t-3}^*} = \sqrt{Sby_{t-3}} \quad \sqrt{Sby_{t-4}^*} = \sqrt{Sby_{t-4}} \\ \sqrt{Sby_{t-1}^*} &= \sqrt{Sby_{t-1}} \quad \sqrt{Sby_{t-2}^*} = \sqrt{Sby_{t-2}} \quad \sqrt{Sum_{t-3}^*} = \sqrt{Sum_{t-3}} \quad \sqrt{Sum_{t-4}^*} = \sqrt{Sum_{t-4}} \\ \sqrt{Sum_{t-1}^*} &= \sqrt{Sum_{t-1}} \quad \sqrt{Sum_{t-2}^*} = \sqrt{Sum_{t-2}} \quad \sqrt{Sby_{t-4}^*} = \sqrt{Sby_{t-4}} \quad \sqrt{Sby_{t-5}^*} = \sqrt{Sby_{t-5}} \\ \sqrt{Sby_{t-2}^*} &= \sqrt{Sby_{t-2}} \quad \sqrt{Sby_{t-3}^*} = \sqrt{Sby_{t-3}} \quad \sqrt{Sum_{t-4}^*} = \sqrt{Sum_{t-4}} \quad \sqrt{Sum_{t-5}^*} = \sqrt{Sum_{t-5}} \end{aligned}$$

And $\varepsilon_{1t}, \varepsilon_{2t}$, were white noise vectors. The result of parameter estimation for transformed and differenced data can be seen at Table 7.

Table 7. Parameter Estimation of VAR (4)

Eq	Parameter	Estimate	Std Error	T Ratio	Prob> T	Variable
Sby(t)	1_1_1	-0.36663	0.07732	-4.74	0.0001	sby(t-1)
	1_1_2	-0.05808	0.06721	-0.86	0.3888	sum(t-1)
	2_1_1	-0.28206	0.0812	-3.47	0.0007	sby(t-2)
	2_1_2	0.05423	0.07057	0.77	0.4433	sum(t-2)
Sum(t)	3_1_1	-0.17095	0.08117	-2.11	0.0367	sby(t-3)
	3_1_2	0.03296	0.07073	0.47	0.6418	sum(t-3)
	4_1_1	-0.06494	0.07742	-0.84	0.4028	sby(t-4)
	4_1_2	-0.0475	0.0675	-0.7	0.4826	sum(t-4)
Sum(t)	1_2_1	0.04506	0.08705	0.52	0.6054	sby(t-1)
	1_2_2	-0.39019	0.07567	-5.16	0.0001	sum(t-1)
	2_2_1	0.03409	0.09142	0.37	0.7097	sby(t-2)
	2_2_2	-0.16938	0.07945	-2.13	0.0345	sum(t-2)
	3_2_1	0.09649	0.09139	1.06	0.2926	sby(t-3)
	3_2_2	-0.21987	0.07963	-2.76	0.0064	sum(t-3)
	4_2_1	0.20176	0.08717	2.31	0.0218	sby(t-4)
	4_2_2	-0.15293	0.07599	-2.01	0.0458	sum(t-4)

Values at Table 7 were still being initial estimated parameter values. Actually not all parameters were significantly different from zero at confidence level 5%, that's why backward elimination should be conducted to get the best model, as shown at Table 8.

Table 8. Final Parameter Estimation of VAR (4)

Eq	Parameter	Estimate	Std Error	t Ratio	Prob> t	Variable
Sby(t)	1_1_1	-0.337	0.074	-4.58	0.000	sby(t-1)
	1_2_1	-0.219	0.074	-2.98	0.003	sby(t-2)

Table 8. Parameter Estimation of VAR (4) (continued)

Sum(t)	Parameter	Estimate	Std Error	t Ratio	Prob> t	Variable
Sum(t)	3_1_3	-0.389	0.075	-5.19	0.000	sum(t-1)
	3_2_3	-0.171	0.079	-2.17	0.032	sum(t-2)
	3_3_3	-0.212	0.079	-2.71	0.008	sum(t-3)
	3_4_1	0.170	0.080	2.11	0.036	sby(t-4)
	3_4_3	-0.153	0.075	-2.04	0.043	sum(t-4)

Table 8 contains final estimated parameters of VAR (4) model resulted from significant parameter. Estimated parameter is said to be significant if Prob>|t| is less than $\alpha=0.05$. Final VAR (4) model of transformed and differenced wind speed data can be written in the form of matrix as below.

$$\begin{aligned} \sqrt{Sby_t} &= \sqrt{Sby_{t-1}} \quad 0.337 \quad 0 \quad 0 \quad \sqrt{Sby_{t-1}} \quad \sqrt{Sby_{t-2}} + \\ \sqrt{Sum_t} &= \sqrt{Sum_{t-1}} \quad 0 \quad 0 \quad 0.389 \quad \sqrt{Sum_{t-1}} \quad \sqrt{Sum_{t-2}} + \\ &0.219 \quad 0 \quad 0 \quad \sqrt{Sby_{t-2}} \quad \sqrt{Sby_{t-3}} + \\ &0 \quad 0 \quad 0.170 \quad \sqrt{Sum_{t-2}} \quad \sqrt{Sum_{t-3}} + \\ &0 \quad 0 \quad 0 \quad \sqrt{Sby_{t-3}} \quad \sqrt{Sby_{t-4}} + \\ &0 \quad 0 \quad 0.212 \quad \sqrt{Sum_{t-3}} \quad \sqrt{Sum_{t-4}} + \\ &0 \quad 0 \quad 0 \quad \sqrt{Sby_{t-4}} \quad \sqrt{Sby_{t-5}} + \epsilon_{1t} \\ &0.170 \quad 0 \quad 0.153 \quad \sqrt{Sum_{t-4}} \quad \sqrt{Sum_{t-5}} + \epsilon_{2t} \end{aligned}$$

The above equation can be written in different form separately for each variable at the time-t

1. $Sby_t = [0.663\sqrt{Sby_{t-1}} + 0.118\sqrt{Sby_{t-2}} + 0.219\sqrt{Sby_{t-3}}]^2$
2. $Sum_t = [0.611\sqrt{Sum_{t-1}} + 0.218\sqrt{Sum_{t-2}} \quad 0.041\sqrt{Sum_{t-3}} +$
 $0.059\sqrt{Sum_{t-4}} + 0.170\sqrt{Sby_{t-4}} \quad 0.170\sqrt{Sby_{t-5}} + 0.153\sqrt{Sum_{t-5}}]^2$

Those model indicates that beside affected by itself at one, two, three, four, and five days before; wind speed in Sumenep also affected by wind speed in Surabaya at four and five days before, while wind speed in Surabaya is only affected by itself at one, two and three days before.

4.5. Diagnostic Checking

4.5.1. Testing of Multinormality Assumption

Multinormality of residual represents one of assumption which must be fulfilled in VAR model. If residual follows multinormal distribution, the data really comes from stationer process. Tool used for examining multinormality is relied on square distance. Testing at $\alpha = 5\%$ was done for hypotheses below:

H_0 : Residual follows multinormal distribution

H_1 : Residual does not follow multinormal distribution

Test statistics:

$$d^2_j = (x \quad x)' S^{-1} (x_j \quad x)$$

(Johnson and Wichern [12])

Based on test statistics d_j^2 , more than a half of d_j^2 values (64.09%) are smaller than $\chi^2_{(2, 50\%)} = 1,39$. Therefore we can not reject null hypothesis or in the other words we can say that residual follows multinormal distribution. This conclusion is also strengthened by *chi-square* plot that seems to be a straight line.

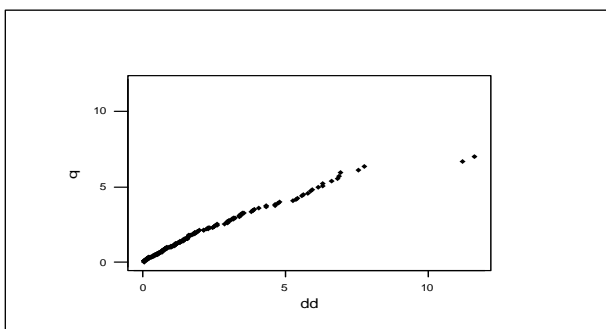


Figure 3. χ^2 multivariate plot of residual

4.5.2. Checking of White Noise Residual

Residuals are said to be *white noise* if residuals follow uncorrelated random series with mean=0 and constant variance. Tool used for checking white noise residual is matrix of residual correlation as follows (Wei [6]):

Hypotheses:

H_0 : $\rho_1 = \rho_2 = \dots = \rho_m = 0$

H_1 : at least one non zero

Test statistics:

$$X(s) = n \sum_{i=1}^m \sum_{j=1}^m [P_{ij}(s)]^2$$

Test statistics values laid on Table 9 show that we can not reject null hypothesis since all X(s) are smaller than $\chi^2_{(5\%,4)} = 9,49$. MACF of residual also suggest the same conclusion that residuals are white noise.

Table 9. Testing of White Noise Residual

Lag	$\chi(s)$
1	1.078
2	1.607
3	5.845
4	2.764
5	5.852
6	1.183
7	1.898
8	8.685
9	3.029
10	3.213

4.5.3. Testing of Homogeneity of Residual Variance

One of another important assumption is homogeneity test of residual variance through Q-test statistic. Testing at $\alpha = 5\%$ was done for hypotheses below:

H_0 : There is no ARCH process in residual of model (Homoskedasticity)

H_1 : There is an ARCH process in residual of model (Heteroskedasticity)

Based on the results from SAS 8, we can not reject null hypothesis for all model since $p\text{-value} > \alpha = 5\%$. This means that there is no ARCH process in residual of model or homogeneity assumption of variance residual was satisfied. Full examination result summarized at Table 10 and 11.

Table 10. Homogeneity Test of Residual Variance for Surabaya

Order	Q	Pr > Q	LM	Pr > LM
1	0.2498	0.6172	0.2769	0.5987
2	0.9501	0.6218	0.8953	0.6391
3	1.1319	0.7694	1.028	0.7945
4	2.1779	0.7031	2.0457	0.7274
5	3.8285	0.5744	3.4273	0.6344

6	3.962	0.6818	3.6035	0.7301
7	4.6542	0.7021	3.9288	0.7879

Table 10. Homogeneity Test of Residual Variance for Surabaya (continued)

8	5.1217	0.7445	4.6893	0.7902
9	5.2829	0.809	4.9949	0.8348
10	5.5878	0.8486	5.2642	0.8728
11	5.7266	0.891	5.2917	0.9162
12	5.7936	0.9261	5.4028	0.9432

Table 11. Homogeneity Test of Residual Variance for Sumenep

Order	Q	Pr > Q	LM	Pr > LM
1	1.5093	0.2192	1.3865	0.239
2	1.8824	0.3902	1.9428	0.3786
3	1.9462	0.5836	1.9568	0.5814
4	2.1183	0.714	2.1547	0.7073
5	2.2594	0.8122	2.2713	0.8105
6	2.612	0.8557	2.7031	0.8451
7	2.683	0.9127	2.7649	0.9059
8	6.1902	0.6259	6.9363	0.5435
9	7.3036	0.6055	7.5042	0.5848
10	8.7713	0.5539	8.6513	0.5655
11	10.4987	0.4862	9.3744	0.5874
12	15.5933	0.2106	14.6659	0.2602

Since all assumptions were satisfied hence the best model was VAR(4) of square root-transformed and differenced data.

5. Conclusions

From data analysis and discussion, it can be concluded that:

1. The best model for wind speed in Surabaya and Sumenep using multivariate time series yields VAR(4) (as shown at page 5).
2. Wind speed in Surabaya is only affected by itself at one, two and three days before.
3. Beside affected by itself at one, two, three, four, and five days before; wind speed in Sumenep also affected by wind speed in Surabaya at four and five days before.
4. The strong relationship between wind speed in Surabaya and Sumenep is also shown by a high measure of correlation between them.

REFERENCES

- [1] Nuryana, F. 2001. Perbandingan Pendekatan ARIMA dan ARFIMA (Studi Kasus Pemodelan Kecepatan Angin di Sumenep). Final Report, Department of Statistics, Institute of Technology Sepuluh Nopember, Surabaya.
- [2] Hartanti, N. 2002. Pemodelan AR(1) dengan Data Hilang (Studi Simulasi dan Studi Kasus Kecepatan Angin di Kota Surabaya). Final Report, Department of Statistics, Institute of Technology Sepuluh Nopember, Surabaya.
- [3] Ewing, B. T., J. B. Kruse, J. L. Schroeder, and D.A. Smith. 2004. Time Series Analysis of Wind Speed Using VAR and the Generalized Impulse Response Technique. Working Paper, forthcoming in *Environmetrics*. The Center for Natural Hazards Research, East Carolina University
- [4] Chatfield, C. 1996. *The Analysis of Time Series An Introduction*, Fifth Edition. Chapman & Hall/CRC, Boca Raton, USA.
- [5] Hamilton, J. D. 1994. *Time Series Analysis*. Princeton University Press, Princeton.
- [6] Wei, W.W.S. 1990. *Time Series Analysis Univariate and Multivariate Methods*. Addison-Walley Publishing Company, Inc., Ontario.
- [7] Jiang, H., dan Kitagawa, G. (ed.) 1999. *The Practice of Time Series Analysis*. Springer, New York.
- [8] Basri, M. C. (2002), "Why Trends of Protection Changed Over Time in Indonesia", *Institute of Asian Studies*
- [9] Enders, W. 1995. *Applied Econometric Time Series*. John Wiley & Sons Inc., Ontario.
- [10] Kapetanios, G. 2002. A Note on Iterative Least Square Estimation Method for ARIMA and VARMA. Technical Report. *Departement of Economics, Queen Mary, University of London*, no. 467.
- [11] Draper, N. dan Smith, H. 1992. *Analisis Regresi Terapan*, edisi Kedua. Gramedia Pustaka Utama, Jakarta.
- [12] Johnson, R.A., dan Wichern, D.W. 1988. *Applied Multivariate Statistical Analysis*. Fourth Edition, Prentice Hall, New Jersey.

Variable Structure Control of Two-Wheels Inverted Pendulum Mobile Robot

Nawawi S.W^{1*}, Ahmad M.N¹ and Osman J.H.S¹

¹ Faculty of Electrical Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5535236, E-mail: sophan@fke.utm.my

Abstract

In this paper the position and posture control for the wheeled inverse pendulum type mobile robot is discussed based on simulation results of two types of controller. The robot in this consideration has two independent driving wheels in same axis, and the gyro type sensor to know the inclination angular velocity of the body and rotary encoders to know wheels rotation. This paper will discuss the control algorithm to make the robot autonomously navigate in two dimensional plane while keeping balance its pole.

Keywords:

Two-wheels inverted pendulum, sliding mode control, nonlinear controller

1. Introduction

Some previous researchers on such a wheeled inverse pendulum type robot have been reported. Wheeled inverse pendulum model have evoked a lot of interest recently [1][2][3] and at least one commercial product (Segway) is available [4]. Such vehicles are of interest because they have a small foot-print and can turn on dime. The kinematic model of the system has been proved to be uncontrollable [3] and therefore balancing of the pendulum is only achieved by considering dynamic effects.

Grasser et al [6] derived a dynamic model using a Newtonian approach and the equation were linearized around an operating point to design a controller. In Solerno et al [3] the dynamic equations were studied, with the pendulum pitch and the rotational angles of the two wheels as the variables of interest. Various controllability properties of the system in terms of the state variables were analyzed using differential-geometric approach. T.Kawamura and K.Yamafuji [5] proposed posture and driving control algorithm of similar vehicle. They assumed that the robot has the tactile sensor to detect posture angle between ground and body. They made a good experiment on balancing but the robot could not move in two dimensional plane because the wheels is driven by single servo motor. O.Matsumoto, S.Kajita and K.Tani [6] presented the estimation and control algorithm of the posture using the adaptive observer. The presented algorithm also did not considered the moving control on the two dimensional plane.

E.Koyanagi et al [7] proposed two dimensional trajectory control algorithm for this type of robot and implemented on the real autonomous self contained robot but the algorithm could only work while the robot moves slowly. Y.Ha and S.Yuta [8] propose the algorithm of trajectory tracking of the wheeled inverse pendulum type mobile robot which can run in relatively high speed in the two dimensional plane.

The focus of this research is to make the wheeled inverse pendulum type robot move smoothly in balancing with proper velocity control. The robot is assumed to have two independent driving wheels in same axis to support and move the robot itself. The gyro sensor is attached to know the inclination angular velocity of the body and rotary encoders to know wheels rotation angle. Based on simulation results which getting from control algorithm of [8], it can be shown that the control algorithm can be improved in terms of robustness by other robust controller. The SMC is used to

$$(M_b r l + M_b l^2 + I_b) \ddot{\phi} + [(M_b + M_w) r^2 + M_b r l + I_w] \ddot{\theta} + \mu_g \dot{\theta} - M_b g \phi = 0$$

compare the simulation result of integral state feedback by [8].

2. Dynamic Model

This system is modeled based on wheels axle and its vertical axis. And body motion in one dimensional plane is determined by the inclination and translation motion. The model of the robot is shown in figure 1. where, θ and ϕ are the wheels rotation angle and the inclination angle of the body respectively and β be the wheel's relative rotation angle to body ($\theta - \phi$). Lagrange motion equation of this model is given as equation (1) and (2) [8].

*Corresponding Author. E-mail: sophan@fke.utm.my

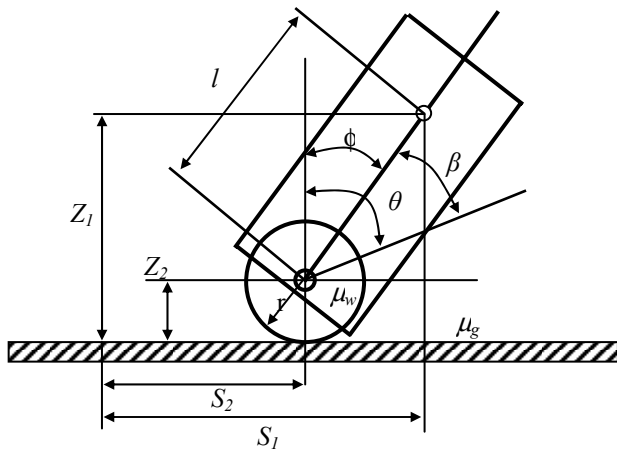


Figure 1: Geometric Parameters and coordinate systems of two wheeled inverted pendulum vehicle

The parameters and variable in equation (8) and (9) is define in Table 1[8].

Value	Parameter and variable name
$M_b = 9.01\text{kg}$	Mass of the body
$M_w = 0.51\text{kg}$	Mass of wheel
$J_b = 0.228\text{kgm}^2$	Rotational inertia of the body
$J_w = 5.1 \times 10^{-4}\text{kgm}^2$	Rotational inertia of the wheel
$J_M = 3.2 \times 10^{-6}\text{kgm}^2$	Rotational inertia of motor axis
$R = 0.062\text{m}$	Radius of the wheel
$l = 0.138\text{m}$	Length between the axle of the wheel and gravitational center of the robot body.
$\mu_w = 0.00576\text{Nm/(rad/s)}$	Viscous between the wheel axle including motor and gear
$\mu_g = 0.00425\text{Nm/(rad/s)}$	Viscous between the wheel and the ground
$\tau_1 = 0.0235\text{Nm/A}$	Torque constant of the motor
$\eta = 39.5$	Reduction ratio of gear

Table 1: Parameter and variable

The state equation of the linearized model is obtained as (10) from (8) and (9).

$$\dot{x} = Ax + Bu \quad (10)$$

Where:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ a_1 & a_3 & a_5 & 0 \\ a_2 & a_4 & a_6 & 0 \end{bmatrix}, B = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, x = [\varphi \quad \dot{\varphi} \quad \dot{\theta}]^T$$

$$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \\ a_5 & a_6 \end{bmatrix} = \begin{bmatrix} \frac{(a_{22} \ a_{12})M_b gl}{\Delta} & \frac{(a_{11} \ a_{21})M_b gl}{\Delta} \\ \frac{\mu_w a_{22}}{\Delta} & \frac{\mu_w a_{21}}{\Delta} \\ \frac{\mu_w a_{22} + \mu_g a_{12}}{\Delta} & \frac{\mu_w a_{21} + \mu_g a_{11}}{\Delta} \end{bmatrix}$$

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \frac{a_{22} \eta \tau_1}{\Delta} \\ \frac{a_{21} \eta \tau_1}{\Delta} \end{bmatrix}, \Delta = (a_{11} a_{22} \ a_{12} a_{21})$$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} M_b l^2 + I_b + \eta^2 I_M & M_b r l \ \eta^2 I_M \\ M_b r l + M_b l^2 + I_b & (M_b + M_w) r^2 + M_b r l + I_w \end{bmatrix}$$

3. Controller Design

In order to synthesize the Sliding mode controller we write the state variable as $x = [\varphi \quad \dot{\varphi} \quad \dot{\theta}]$ and get the following state-space matrices[9]:

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{\beta}} - \frac{\partial T}{\partial \beta} + \frac{\partial U}{\partial \beta} + \frac{\partial D}{\partial \dot{\beta}} = Q_\beta \quad (1)$$

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{\theta}} - \frac{\partial T}{\partial \theta} + \frac{\partial U}{\partial \theta} + \frac{\partial D}{\partial \dot{\theta}} = Q_\theta \quad (2)$$

Where:

T=Kinetic energy

U=Potential energy

D=Dissipation energy function

Q_β =External force to β axis

Q_θ =External force to θ axis

$$T = \frac{1}{2} M_w (\dot{s}_2^2 + \dot{z}_2^2) + \frac{1}{2} M_b (\dot{s}_1^2 + \dot{z}_1^2) + \frac{1}{2} I_w \dot{\theta}^2 + \frac{1}{2} I_b (\dot{\theta} \ \dot{\beta})^2 + \frac{1}{2} I_M \eta^2 \dot{\beta}^2 \quad (3)$$

$$U = M_w g r + M_b g l \cos(\theta \ \beta) \quad (4)$$

$$D = \frac{1}{2} (\mu_w \dot{\beta}^2 + \mu_g \dot{\theta}^2) \quad (5)$$

$$Q_\beta = \eta \tau_1 u_1 \quad (6)$$

$$Q_\theta = 0 \quad (7)$$

Assume the $\varphi = 0, \dot{\varphi} = 0$ to linearized in the neighborhood of the up right state.

$$(M_b l^2 + I_b + \eta^2 I_M) \ddot{\varphi} + (M_b r l \ \eta^2 I_M) \ddot{\theta} + \mu_w \dot{\varphi} \ \mu_w \dot{\theta} \ M_b g l \varphi = \eta \tau_1 u_1 \quad (8)$$

$$(M_b r l + M_b l^2 + I_b) \ddot{\varphi} + [(M_b + M_w) r^2 + M_b r l + I_w] \ddot{\theta} + \mu_g \dot{\theta} \ M_b g l \varphi = 0 \quad (9)$$

$$A = \begin{bmatrix} 0 & 1 & 0 \\ a_1 & a_3 & a_5 \\ a_2 & a_4 & a_6 \end{bmatrix}, B = \begin{bmatrix} 0 \\ b_1 \\ b_2 \end{bmatrix}, x = [\varphi \quad \dot{\varphi} \quad \dot{\theta}]^T$$

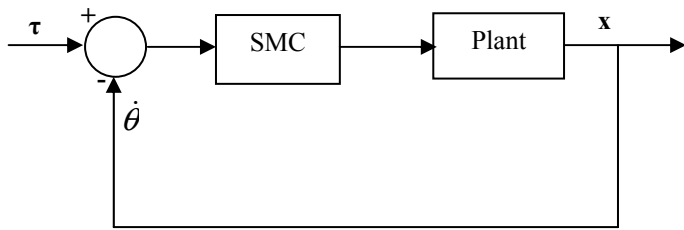


Figure 2: Block diagram of SMC for self-contained mobile inverse pendulum

To develop a sliding mode controller, equation (10) has been decompose as:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u \quad (11)$$

$$\begin{aligned} x_1 &= \varphi \\ x_2 &= \dot{\theta} \end{aligned}$$

let the switching surface be defined as $\sigma = Sx$ where

$$\sigma = \begin{bmatrix} S_1 & S_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, S_1 \in R^{1 \times 3}; S_2 \in R \quad (12)$$

assume the system dynamics is an ideal sliding surface, $\sigma = Sx = 0$. Using this property it can be determine the equivalent system and associated linear control input. To proceed, using $\sigma = 0$

$$x_2 = \frac{S_1}{S_2} x_1 \quad (13)$$

Substitute (13) into (11) yields the equivalent system:

$$\dot{x} = \begin{bmatrix} A_{11} & A_{12} S_2^{-1} S_1 \end{bmatrix} x_1 \quad (14)$$

defining

$$\dot{x}_1 = \begin{bmatrix} A_{11} & A_{12} k \end{bmatrix} x_1 \quad (15)$$

the location of poles of the resulting system are obtained by selecting k and S_2 as the switching function becomes $\sigma = Sx, S = \begin{bmatrix} S_1 & S_2 \end{bmatrix} = \begin{bmatrix} S_2 k & S_2 \end{bmatrix} = S_2 \begin{bmatrix} k & I \end{bmatrix}$. Pole placement method is use to select the gain k.

The sliding mode control inputs are separated into the linear and nonlinear component as $u_{eq} = u_l + u_{nl}$. The linear input u_l can be selected by the following equations:

$$\dot{\sigma} = S\dot{x} = 0 \quad (16)$$

from (16) and (10), the equivalent linear control input is:

$$u_l = (SB)^{-1} SAx \quad (17)$$

the sliding-mode reaching condition given by $\sigma\dot{\sigma} < 0$, bring the system dynamics to the sliding surface $\sigma=0$. Choose the nonlinear control a $u_{nl} = (SB)^{-1} \rho \text{sgn}(\sigma)$ where $\rho > 0$. Then:

$$\begin{aligned} \sigma\dot{\sigma} &= \sigma S(Ax + Bu) \\ &= \sigma SAx - \sigma SB \left[(SB)^{-1} SAx + (SB)^{-1} \rho \text{sgn}(\sigma) \right] \\ &= -\rho \sigma \text{sgn}(\sigma) \end{aligned} \quad (18)$$

that is $\sigma\dot{\sigma} < 0$ whenever $\sigma \neq 0$, which means the reaching condition is satisfied. The control input can then be written as:

$$u_{eq} = (SB)^{-1} \left[SAx + \rho \text{sgn}(\sigma) \right] \quad (19)$$

3. Stability Analysis

For a given control system stability is usually the most important attribute to be determine. The stability of this system with SMC is proved to be stable at the designed sliding surface using lyapunov method. The concept is introduced by Russian mathematician A.M Lyapunov. The lyapunov's method of stability analysis is in principle the most general method for determination of stability of nonlinear or time varying systems.

lyapunov's Theorem:

1. $V(t)$ is continuous first partial derivatives
2. $V(t)$ is positive definite
3. $\dot{V}(t)$ is locally negative definite

from Ricati equation :

$$Q = PA + A^T P \quad (20)$$

where matrix Q and P is positive definite.

For linear system time-invariant systems, i.e $f(x,t)=Ax$ a suitable Lyapunov function is $V(x) = x^T Px$ where P is obtain from Ricati equation and Q is a positive definite symmetric real matrix. The control proposed by M.J Corless and G.Lietmann[10] is $u_2 = p(x,t)$ where p is any continuous function which satisfies:

any continuous function which satisfies:

$$p(x,t) = \begin{cases} \frac{B^T Px}{\|B^T Px\|} \rho(x,t) \Rightarrow \|B^T Px\| \rho(x,t) > \varepsilon \\ \frac{B^T Px}{\varepsilon} \rho(x,t)^2 \Rightarrow \|B^T Px\| \rho(x,t) > \varepsilon \end{cases} \quad (21)$$

The proposed control action is in fact a continuous approximation of the discontinuous min-max control. For simplicity the case of n-order single input linear time invariant system without uncertain elements is considered. The model of a linear system can be written as follows:

$$\dot{x} = Ax + Bu$$

The switching hyperplane were define as follow

$$\Sigma = \{x | \sigma(x) = 0 \text{ where } \sigma = Sx \text{ and } S \in \mathbb{R}^{1 \times n} \quad (22)$$

to provide the sliding mode, all the phase trajectories in the vicinity of Σ must be oriented toward it[11]. There exists a lyapunov function $V(x, t, \sigma)$ such that and in $V(x, t, \sigma) = 0$ neighborhood of Σ , the condition $V < 0$ is satisfied. In what follows, the switching surface is assume to be chosen so that the restriction of the nominal system to the surface is asymptotically stable. For the single input case a suitable lyapunov function is:

$$V(x, t) = \frac{1}{2} \sigma^2(x) \quad (23)$$

Thus if $\dot{\sigma} = S\dot{x} = SAx + SBu$ $\dot{V} = \sigma\dot{\sigma} < 0$ in a neighborhood of Σ for all the time, then all state trajectories initially in this neighborhood, converge to the surface and are restricted to the surface for all subsequent time.

$$(24)$$

Assume $SB \neq 0$ and since the hyperplane Σ does not change if S is multiplied by an arbitrary constant, without loss of generality, it can be assume that $SB < 0$. Thus to ensure sliding mode behavior, u can be a any function fulfilling the inequalities:

$$\begin{aligned} u < (SB)^{-1} S(Ax) &\Rightarrow \sigma < 0 \\ u^+ > (SB)^{-1} S(Ax) &\Rightarrow \sigma > 0 \end{aligned} \quad (25)$$

Now since in sliding mode, the state does not leave the reduced order subspace Σ , the state velocity also belongs to Σ , that is, $S\dot{x} = 0$. By the substitution of \dot{x} from (3) the following is obtained:

$$SAx + SBu_{eq} = 0 \quad (26)$$

Since $SB \neq 0$, u_{eq} is determined uniquely as:

$$u_{eq} = (SB)^{-1} S(Ax) \quad (27)$$

The sliding mode equation in Σ is:

$$\begin{aligned} \dot{x} &= [I \quad B(SB)^{-1} S] Ax \\ Sx &= 0 \end{aligned} \quad (28)$$

So for SMC controller:

$$V(t) = x^T P x \quad (29)$$

The lyapunov function $V(t)$ is positive definite since P is positive definite. The derivative of the lyapunov function with respect to time t can be obtained as follows:

$$\dot{V}(t) = 2x^T P \dot{x} \quad (30)$$

At sliding surface:

$$\dot{x} = Ax + Bu_{eq}$$

$$u_{eq} = (SB)^{-1} SAx$$

substitute into \dot{x}

$$\dot{x} = Ax + B \left((SB)^{-1} SAx \right)$$

$$\dot{x} = \left(I \quad B(SB)^{-1} S \right) Ax$$

$$\text{let } \bar{A} = \left(I \quad B(SB)^{-1} S \right) A$$

$$\dot{V}(t) = 2x^T P \dot{x}$$

$$\dot{V}(t) = 2x^T P \bar{A} x$$

$$\dot{V}(t) = x^T [P \bar{A} + \bar{A}^T P] x,$$

substitute Riccati Equation into $\dot{V}(t)$

$$\dot{V}(t) = -x^T Q x$$

From Rayleigh's principle:

$$\begin{aligned} \lambda_{min}(Q) \|x\|^2 &\leq -x^T Q x \leq \lambda_{max}(Q) \|x\|^2 \\ \therefore \dot{V}(t) &= -x^T Q x \leq \lambda_{min}(Q) \|x\|^2 \end{aligned}$$

The derivative of the lyapunov is always negative, which prove via Lyapunov stability theorem that this system equilibrium state at the origin is asymptotically stable. In SMC beside proving the stability at sliding condition, another important thing to prove is the stability of the reaching condition. So for the reaching condition:

$$u_{eq} = (SB)^{-1} [SAx + \rho \operatorname{sgn}(\sigma)]$$

$$u_{eq} = (SB)^{-1} SAx + E$$

$$\text{where: } E = \rho \operatorname{sgn}(\sigma)$$

now for the case of multi input system, assume:

$$\|E\| = \hat{\rho}(x, t)$$

let

$$u^* = (SB)^{-1} SAx + \frac{B^T S^T \sigma}{\|B^T S^T \sigma\|} \rho(x, t) \Rightarrow \|\sigma\| > 0$$

$$(SB)^{-1} SAx \Rightarrow \|\sigma\| = 0$$

And $\rho(x,t) = \alpha + \hat{\rho}(x,t)$, $\alpha > 0$. For such u^* the derivative \dot{V} becomes:

$$\dot{V}(t) = \sigma^T \dot{\sigma} = \sigma^T \left[SAx + SB \left(-(SB)^{-1} SAx - \frac{B^T S^T \sigma}{\|B^T S^T \sigma\|} \rho(x,t) \right) + SBE \right] \leq -\alpha \|B^T S^T \sigma\|$$

It can be seen that the implementation of u^* assures that the derivative of Lyapunov function is negative everywhere outside the switching surface. From the above derivation it can be shown that the SMC designed for the inverse pendulum mobile robot plant is proved to be asymptotically stable.

3. Simulation Result

For the propose controller of SMC, the simulation result can be seen much better then SFIC which propose by Yuta S. [8] in term of time response and percent of overshoot. The sliding surface which using in SMC for the propose controller is $S=[12; 4.6; 1]$. The value of S is tune by heuristic.

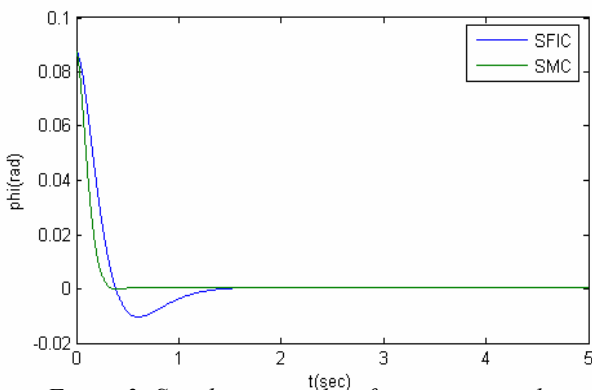


Figure 3: Simulation results of posture control using State Feedback Integral Control (SFIC) and SMC

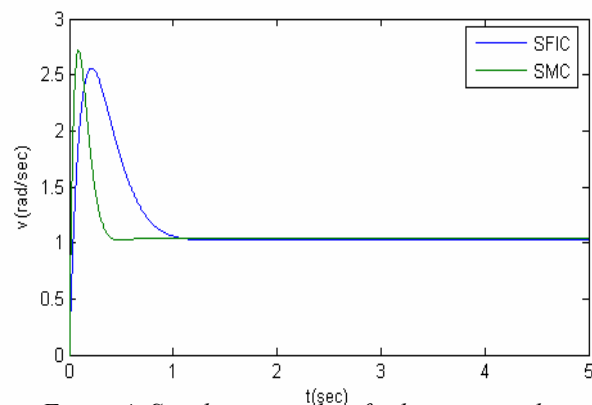


Figure 4: Simulation results of velocity control using State Feedback Integral Control (SFIC) and SMC

4. Conclusion

In this paper a two-wheeled inverted pendulum type robot is discussed. It has the advantage of mobility from without caster and an innate clumsy motion for balancing. The proposed controller in particular SMC shows the good control behavior of the system to be control. It can be seen in the result of the simulation. The characteristics of the SMC appear to be more robust and good control compare to SFIC on inverse pendulum type mobile robot.

References

1. Grasser F., D'Arrigo, A., Colombi, S. and Rufer, A. "Joe; A Mobile Inverted Pendulum", IEEE Trans. Industrial Electronics, vol.49, no 1, pp. 107, no 114, 2002
2. Baloh, M and Parent, M., "modeling and model verification of an intelligent self balancing two-wheeled vehicle for an autonomous urban transportation system". Conf. Comp. Intelligence, Robotics and autonomous systems, Singapore, Dec, 15 2003.
3. Salerno A and Angeles J " On the nonlinear controllability of a quasiholonomic mobile robot", Proc, IEEE ICRA Taiwan 2003.
4. "Segway Human Transporter", <http://www.segway.com> 2003.
5. Yamafuji K. and Kawamura T., "Postural control of monoaxial bicycle", Journal of the robotic society of japan, vol.7 No 4 pp 74-79, 1988
6. Matsumoto O., Kajita S. and Tani K., "Attitude estimation of the wheeled inverted pendulum using adaptive observer", Proc. Of 9th Academic Conf of robotic Society of Japan, pp.909-910, 1991
7. Koyanagi K., Iida S., Kimoto K. and Yuta S., "A wheeled inverse pendulum type self-contained mobile robot and its two dimensional trajectory control" Proc of ICMCR 92, pp.891-898, 1992.
8. Ha, Y.S and Yuta S., "Trajectory tracking control for navigation of the inverse pendulum type self-contained mobile robot", Robotics and autonomous systems, 17, pp. 65-80, 1996.
9. Lee J.H., Allaire P.E., Tao G. and Zhang X., "Integral Sliding mode control of a Magnetically Suspended Balance beam: analysis, simulation and experiment", IEEE/ASME Trans. On mechatronics Vol 6 no 3, September 2001.

Starfruit Grading Based on 2-Dimensional Color Map

M. M. Mokji¹, S.A.R. Abu Bakar²

*Faculty of Electrical Engineering
University of Technology Malaysia
Johor Bahru, Johor, Malaysia
Email: ¹musa@fke.utm.my, ²syed@fke.utm.my*

Abstract

This paper present a grading technique on starfruit based on image processing. 2-dimensional color map is used in order to classify the starfruit into its ripe index. Referring to standard introduced by FAMA, there are six ripe indexes to code in. The color map is based on 8-bit RGB color space where blue component is discarded to simplify the computation process. This paper also shows that calculating the average chromaticity for the 2-dimensional color map (RG map) can determine the most concentrated color in each of the ripe index. The chromaticity value is then used to segment the starfruit image to reduce the amount of the color. From the segmented image, the starfruit is classified into its ripe index.

Keywords:

Color map, color segmentation, ripe index.

1. Introduction

Agriculture products are being more demanding in market today. To increase its productivity, automation to produce these products will be very helpful. One of the growing agriculture products is starfruit. It can be observed that the export of starfruits has been increasing steadily from 6,300 metric ton in 1988 to 18,100 metric ton in 1999. Some of the major export markets for starfruits are EEC such as United Kingdom, Holland and Germany, which takes up about 57% of the total export, Singapore 39% and with the remainder going to Middle East especially Saudi Arabia, United Arab Emirates and Kuwait, Hong Kong, China and Brunei.

Quality of the starfruit is described by its physical appearance and taste. Auspicious to have good taste of the starfruit is worthless if the physical appearance of the starfruit is ignored. However, until this paper is written, the quality inspection of the starfruit was performed manually wherein manual inspection will caused inconsistency in quality due to the human subjective of nature, slow processing and labor intensive. Currently, in the starfruit pack house, the manual grading process involves removal of the damaged starfruit and sorting the starfruit into categories (based on ripe index). Removal of the damage starfruit is done by inspecting the starfruit skin surface defects. Only starfruit with less than 5% defects will go through the next process, which is the classification of the starfruit into six indexes base on the starfruit ripeness. This shows that manual inspection is a

tedious and complicated process. Automation of the process will solve the problems. With automation process, human can improve their quality of work by concentrating their works to other scopes like market planning and how to improve the effectiveness of the grading process.

Actually, automation for fruit grading has been done to apples [1], orange [2], papaya [3] and a few other fruits. However, each fruit has different criteria, which make automation for fruit grading using general machine is impossible. Hence, a specific machine for starfruit grading needs to be designed. Designing automation solution for starfruit is more difficult compared to the other fruit as it has a complicated shape. This is because the starfruit shape is unique as it has five ridges forming a star shape while the other fruit only have a flat surface. Due to this unique shape, features extraction process becomes more complicated and challenging where lighting exposure is not homogeneous to its surface. Another issue is the shiny surface on the starfruit skin. It reflects a high concentration of light and results high intensity color image, which degrades most of the chromaticity information. This means that recognizing color at this area is tough.

2. Image Acquisition

In image acquisition, a web camera is used to capture the starfruit image. Although web camera is not as good as the other sophisticated camera like a CCD camera, its low-priced wins among the other advantages. Besides, the quality of the captured image can be improved by applying good lighting system. In this work, the starfruit is placed

on a conveyor that moving towards a tunnel where the image of the starfruit will be captured. In this tunnel, the normal environment lighting is blocked and a more suitable lighting source is created in the tunnel.

Captured image will be analyzed using a software develop in C language. Categorizing the starfruit into its ripe index consists of three main steps: preprocessing, color segmentation and classification. Preprocessing is to eliminate small structure in the image. This is done by filtering the image using median filter. Other filters that can be used are averaging filter and Gaussian filter. These are actually low-pass filter. Median filter is chosen because it can keep the color boundary of the image, which will alleviate the color segmentation process.

3. Color Segmentation

Color segmentation is important in many computer vision and image processing application. It is to split image into certain division, which has similarity in color information. In general, current color segmentation technique can be roughly classified based on two properties: discontinuity and similarity. Methods that are based on discontinuity property are called boundary-based method and their objective is to extract borders between regions in an image. Whereas, methods based on similarity property is called region-based method. This method will try to partition the image into a number of regions such that each region has the most similarities [4]. These methods are applied in [5], [6], [7], [8] and many others.

In this work, color in starfruit image is segmented into six regions based on chromaticity in 2-dimensional color map. Here, the 2-dimensional color map is referring to RG color map, which is used in this work to replace the typical 3-dimensional color map of RGB. Hence, less and simpler computation can be achieved as the dimension is reduced. The 2-dimensional color map was formed by discarding blue component from the RGB color map. Blue is discarded, as it does not reflect much in starfruit image. Figure 1 shows color component in a starfruit image, where most of the blue component has a small values. By ignoring the blue value (suppress the value to zero), the appearance of the original image will not differ significantly.

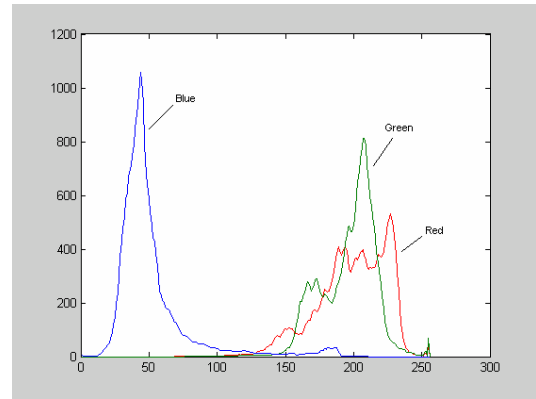


Figure 1. RGB color component.

From the 2-dimensional color map, chromaticity that represents the variation of colors can be computed as below.

$$C = \frac{R}{R + G} \quad (1)$$

Based on Equation 1, six points of chromaticity is computed to be the reference for the color segmentation process. The reason of choosing six points of chromaticity is because the starfruit will be classified into six ripe indexes. Specifically, these points is computed by taking the average chromaticity of each starfruit index in the whole database, which is denoted by

$$I_p = \frac{1}{N} \sum_{q=1}^N C_{pq} \quad (2)$$

In Equation 2, p and q is referring to starfruit index and starfruit number while N is referring to the total number of starfruit in the database of that starfruit index. As p is the starfruit index, it is range from 1 to 6. Having the six values of I , each starfruit image is color segmented according to Equation 3

$$f(i) = I_p \quad (3)$$

where

$$p = 1 : 6 \rightarrow \min\{|C_i - I_p|\} \quad (4)$$

Here, f is the starfruit image and i is the pixel number. Thus, each pixel in the starfruit image is segmented to the nearest chromaticity distance among the six chromaticity values (I_p), which have been computed earlier by Equation 2. However, not all regions will appear in each starfruit image. For an example, starfruit of index 1 most probably will only has region 1 to 3 only and index 6 starfruit will has most of its color in region 4 to 6 only. As

the color of the starfruit image has been segmented and at the same time reduced to only six colors, the next process is to classify the starfruit.

4. Classification

The starfruit is classified into six ripe indexes, which carry out a standard through a label named Malaysia's Best introduced by FAMA (Federal Agricultural Marketing Authority). Index of ripeness is used to determine whether the starfruit is suitable for export purpose or for domestic market. From Index 1 to Index 6, the ripeness of the starfruit change from an immature to mature fruit. For export purpose, only Index 2, Index 3 and Index 4 are allowed. Exporting immature starfruit is to ensure that the starfruit will only be matured at the time it arrive its destination. For domestic market, Index 5 and Index 6 are the most suitable as it can be eaten at the time the fruit is bought by consumer.

From the segmented image discussed in the previous topic, the starfruit is classified into the ripe index based on the amount of pixel in each of the segmented region. The rules are described as in the Figure 2 where R is a short form for region.

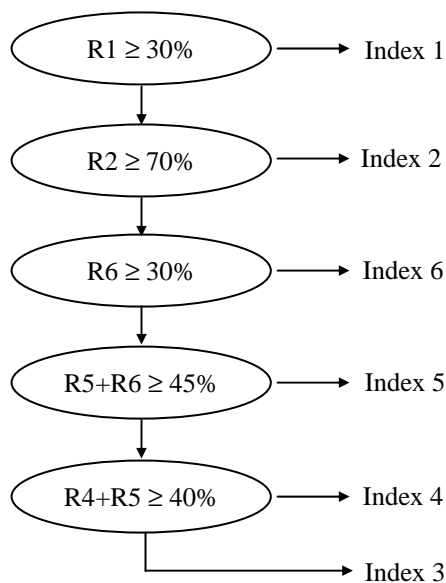


Figure 2. Classification rules

5. Experimental Results

Technique used in this work has been tested on 100 samples of starfruits. Each starfruit index contained 15 samples and there are 6 indexes all together. The other 10 samples are picked among the defected starfruit with various indexes. There are 10 types of defect on starfruit. Each and every single of the defect can be recognized by the physical appearance of the starfruit. This paper is not going to explain the detail about those defects as the main

concern of the paper is on classifying the starfruit into the ripe index.

Firstly, samples are sort by human expert into its ripe index. Then it is tested using technique presented in this paper. The result produced 96% of the starfruit were correctly classified into the ripe index. Table 1 shows the results for each of the ripe index where incorrect classifications are 1% from the defected fruit and 3% from Index 6. Most of the misclassification is due to the light reflection on the starfruit shiny skin, which had degraded most of the color information. Degraded means that the skin colors of the starfruit becomes brighter toward the pure white color. This situation is less occurred on the starfruit with Index 1 to Index 4. This is the reason why they results 100% of accuracy. Thus, proper lighting condition should be applied to solve the shiny skin problem for Index 1 to Index 4.

Table 1. Classification Results

Ripe Index	Accuracy
Index 1	100%
Index 2	100%
Index 3	99%
Index 4	95%
Index 5	100%
Index 6	80%
Defected	90%
Average	94.9%

6. Conclusion

The presented work has shown that classifying the fruit into its ripe index can be accomplished using two-dimensional color map rather than the three-dimensional color map. Here starfruit is use as the subject. The technique plots magnitude of the red and green component on RG color map and calculate the chromaticity of the plot to determine the ripe index. After tested on 100 samples, it produced a result with only 5.1% of the samples are wrongly classified. Most of the misclassification is due to the high light reflection on the skin surface of the starfruit. A better lighting source should increase the classification accuracy. And it is believe that the technique will work on any fruit that has the same behavior of ripening as the starfruit does.

References

- [1] Q. Yang, "Automatic detection of patch-like defects on apples", Fifth International Conference on Image Processing and its Applications, 1995, Page(s):529-533, 4-6 Jul 1995
- [2] M. Recce, J. Taylor, A. Piebe, G. Tropiano, "High speed vision-based quality grading of oranges" International Workshop on Neural Networks for Identification, Control, Robotics, and Signal/Image Processing, 1996, Page(s):136-144, 21-23 Aug. 1996

- [3] S. Limsiroratana, Y. Ikeda, "On image analysis for harvesting tropical fruits", Proceedings of the 41st SICE Annual Conference, Volume 2, Page(s): 1336-1341 vol.2, 5-7 Aug. 2002
- [4] X. Munoz, J. Freixenet, X. Cufi, J. Marti, "Active regions for colour texture segmentation integrating region and boundary information", 2003 International Conference on Image Processing, Volume 3, 14-17, Page(s):III - 453-6 vol.2, Sept. 2003.
- [5] M. Ruzon and C. Tomasi, "Color Edge Detection with the Compass Operator", Proc. Of IEEE Conf. On CVPR, volume 2, Page(s): 160-166, June 1999
- [6] W. Y. Ma and B. S. Manjunath, "Edge Flow: A Framework of Boundary Detection and Image Segmentation", Proc. Of IEEE Conf.on CVPR, Page(s): 750-755, June 1997
- [7] D. Comaniciu and P. Meer, "Robust Analysis of Feature Space: Color Image Segmentation", Proc. Of IEEE Conf. On CVPR, Page(s): 750-755, June 1997
- [8] J. Shi and J. Malik, "Normalized cuts and Image Segmentation", IEEE Trans. On PAMI, Vol. 22, No. 8, Aug 2000

Overview of a Computer-based Stuttering Therapy

Ooi Chia Ai ¹, J. Yunus ²

¹ Department of Electronic Engineering
Faculty of Electrical Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +6012-5844183, E-mail: OoiChiaAi@yahoo.com

² Department of Electronic Engineering
Faculty of Electrical Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5535207, E-mail: jasmy@fke.utm.my

Abstract

The purpose of this paper is to introduce the reader to a variety of anti-stuttering devices that have been used to treat stuttering. The use of devices to cure, or at least control, stuttering is not new. A variety of devices have been utilized, with a variety of purposes. A few examples of each of the different classes of devices are discussed. Reviews on these various electronic devices are presented where there is no one best treatment for stuttering for every client. It is suggested that choosing different types of stuttering managements is a personal decision based on the unique values and treatment objectives of each individual person who stutters. This leads to the implementation of our computer-based Malay stuttering therapy to help Speech-Language Pathologist (SLP) to determine suitable technique for each client. The project implements Digital Signal Processing (DSP) techniques to analyze speech signals and incorporates fluency shaping techniques. Our hypothesis is by doing comparison between client's and clinician's Average Magnitude Profiles (AMPs) for different techniques, the computer analysis can help SLP to determine suitable techniques in faster and more accurate manner. This will reduce cost and time for therapy. We believe our software tool will improve the effectiveness and availability of stuttering therapy in Malaysia. We hope that our software tool will provide insights into the implementation of computer-based Malay stuttering therapy in Malaysia.

Keywords:

Stuttering, Speech Fluency, Computer-Based, Therapy Techniques, Electronic Devices

1. Introduction

Stuttering is one of the serious problems in speech pathology. It occurs in 1% of populations and has found to affect four times as many males as females. Stuttering [1] is the disorders in the rhythm of speech in which the individual knows precisely what he wishes to say, but at the time is unable to say it because of an involuntary, repetitive prolongations or cessation of a sound.

There are basically four classes of computer-based stuttering devices available on the market. Some devices are aimed at forcing the person who stutters to change the manner in which they speak. Others are utilized to provide the client with more easily observable feedback of their current physiologic status or production pattern, often to assist them in achieving speech production targets utilized in therapy. Still others alter the natural feedback the client receives in an attempt to prevent stuttering. We will discuss one by one in the following sections.

2. Devices that Alter Auditory Feedback (AAF)

There are four types of AAF [2]. They are Masked Auditory Feedback (MAF), Delayed Auditory Feedback (DAF), Frequency Altered Feedback (FAF) and Combined/multiple feedback (DAF, FAF, MAF). The effectiveness and value of auditory feedback has been documented in the professional literature [3]. Indeed, portable devices are available for implementing auditory feedback. Most of these devices are single purpose especially DAF devices. Some of the feedback devices are intended to disrupt feedback (DAF and MAF), and others are intended to enhance normal auditory feedback. MAF refers to the use of sound, generally a noise of some sort, of sufficient intensity to block the auditory feedback of the speaker's own voice to his or her ears. With DAF [4], we hear what we say a short period of time after we say it. The stutterer talks into a microphone and his or her speech is recorded and played back through speakers or headphones at a delay of 250 milliseconds [5]. Some stutterers stutter less severely when they experience DAF while they speak [6]. The most recent of the AAF modes to be introduced is

frequency altered feedback, also known as frequency shifted feedback (FSF). The effects of FAF in stutterers have been examined in a variety of paradigms, including frequency shifting in either direction, at various speaking rates [7, 8].

Pocket DAF/FAF Assistant [9] is a software application that implements DAF and FAF techniques on handheld computers running Windows Mobile 2003 for Pocket PC and Windows Mobile 5.0 operating system. The application can be used by people having stuttering problem to control their speech fluency, increase their confidence level and develop the carryover fluency when the techniques are used on a regular basis. Pocket DAF/FAF Assistant delays the voice to ears a fraction of a second later. The application provides delay ranged from 50 to 250 milliseconds. The delay increment step is 10 milliseconds. A short delay (50-80 milliseconds) almost instantly reduces stuttering without changing the speech speaking rate. A longer delay (90-250 milliseconds) enables stutters to stretch vowels and talk slower. It can help with even severe stuttering, but requires training, and mental effort. FAF shifts the pitch of voice and provides the pitch shift in the range from one-half octave down to one-half octave up. The FAF enhances the effectiveness of the application when it is used simultaneously with DAF [10].

The SpeechEasy [11] is digitally programmable with special computer software for a wide range of settings of DAF and FAF. SpeechEasy fluency devices are based on a natural phenomenon called the "choral effect." The "choral effect" occurs when people who stutter speak or sing in unison with others and their stuttering is dramatically reduced or even eliminated. SpeechEasy employs AAF in the form of auditory delays and frequency shifts to provide maximum long-term benefit to individuals who stutter. They are adjusted to fit in or behind the ear, and are to be worn like traditional hearing aids. Initial clinical trials of SpeechEasy have been conducted at the Stuttering Research Laboratory, East Carolina University. This device has demonstrated a 80-90% success rate in the treatment of over 200 patients of all ages who suffer from fluency disorders, with levels of fluency enhancement varying from 50% improvement up to 95% improvement [11].

Auditory Feedback Tools (AFT) [12] is a program option which provides five different auditory feedback tools in a software package for therapy and feedback applications. The AFT program provides DAF, as a form of disruptive feedback, which has proven to be effective in fluency therapy. The DAF in AFT has a range of feedback from 150-500 milliseconds, adjustable in 10-millisecond increments. In masking mode, a speech-band noise signal is played through headphones so that patients cannot hear their own speech production. This deliberately degraded feedback has been shown, in some cases, to improve speech [12]. In many patients, it can enhance the proprioception of speech or voice behaviors such as easy onset and eliminating hard glottal attack.

Transcutaneous Electrical Nerve Stimulation (TENS) consists of the pocket sized processing unit, a lapel microphone, self-adhesive gel pad TENS electrodes, and a set of headphones which can be used to provide an optional DAF. Speech is picked up by the lapel microphone, delayed

from 15ms to 150ms, amplified, and delivered to the TENS electrodes. The electrodes provide a mild electrical shock to the skin, in the form of a "tingling" sensation, whose strength can be adjusted. According to Dr. Shames [13], the use of TENS feedback results in less adaptation and hence greater awareness of the use of sustained phonation.

3. Devices that Provides Feedback on Physiological Status or Production Patterns

Several therapy programs utilize electronic analyses and feedback of acoustic speech characteristics. These devices provide immediate feedback of voice onset patterns, duration, and amplitude/loudness. As each speech production target is introduced to the clients, they receive training on the type of feedback from the voice monitor that indicates the correct achievement of the target. Once the clinician has confirmed that the clients know what to look for on the voice monitor, the clients could practice without the clinician's presence, freeing up the clinician while the clients engage in intensive practice on their targets.

FluencyNet [14] uses personal computer-based hardware and software to replace the voice monitor. The software program also presents the stimuli (sounds, words, and longer utterances) to practice for each target, data on performance over time, and an opportunity for the client to self-judge the accuracy of their production before feedback is given. FluencyNet is based upon the physical analysis of speech sounds as they are being uttered. FluencyNet provides real-time measures of sounds, evaluate the sounds against standards for their production, and immediately signal the results of the evaluation in graphs plotted on the computer screen. When the onscreen graph shows an accumulation of many green bars and few red bars, the speech signals are judged to be mostly correct. On the other hand, when the onscreen graph shows an accumulation of many red bars and few green bars, the speech signals are judged to be outside the limits necessary for the maintenance of fluent speech.

Digital Speech Aid (DSA) [15] is designed based on an advanced DSdP (Digital Sound Processing) of speech signal in the auditory feedback loop. The device uses the most modern DSP circuitry and it is designed according to the algorithms developed to elicit the desired response characteristics of the feedback signal. All known and many new algorithms can be easily implemented in the existing hardware, due to the fact that DSdP software is stored in EPROM, which could be easily changed and reprogrammed. With DSA, a person can speak in any fashion and at any rate. DSA is most effective in the case of "Classical Stutterers" who consist about 80% - 90% of the stuttering population. Significant improvement or total fluency is observed in about 40% - 60% of "Classical Stutterers".

4. Devices that Alter Speech Motor Production Patterns

Fluency Master [16] is a miniature, wearable, electronic stuttering control device that looks like a hearing

aid. Research [16] has shown that the speech muscles of stutters do not perform the correct sequence of movements necessary to produce fluent speech. The Fluency Master reduces stuttering by modifying physical factors that affect speech, giving the brain the ability to more effectively control the movement of speech muscles. Fluency Master fosters better hearing of the natural vocal tone associated with speaking. Vocal tone provides information for the brain to use in guiding speech muscle movements. Vocal tone consists of a "buzz" that is generated by vibrations of the vocal folds in the larynx, and is transferred to the ears through the soft tissues, cartilage, and bones of the throat and skull. Vocal tone is a natural, internal component of speech, and provides a "background" sound that is always present inside our bodies when we talk.

Stuttering controlled by Fluency Master involves the use of a miniature microphone to detect vocal tone vibrations. The vocal signal is amplified and processed by electronic circuitry housed in a small hearing aid case. A tiny plastic tube carries the sound to a custom-fitted ear-piece.

5. Pacing/Metronome

Many stutters stutter less frequently when they pace their speech while reading aloud or doing a spontaneous speech task with the beats of a metronome [17]. Doing so will cause stutters to concentrate on how they are speaking and thus reduce their speaking rates. This technique has been used clinically for several centuries. The metronome beat can be delivered auditorily, visually, tactilely, or by some combination of these senses. The client is told to pace his or her speech while reading aloud or doing a spontaneous speech task with the beats of a metronome which is one word per beat.

Researchers, clinicians, and people who stutter have long been aware that speaking to a superimposed rhythm can facilitate a noticeably more fluent speaking pattern in many individuals who stutter. AFT [12] provides metronomic pacing in the form of an audible click to help assist patients with the timing/rhythm aspects of speech production. Among these are stutters and patients with motor speech disorders such as cerebellar ataxia and Parkinson's disease. The pacer rate is adjustable from 50 to 150 beats per minute adjustable in 5 beat increments.

6. Computer-Based Malay Stuttering Therapy

The use of computer technology in speech therapy and assessment is still new in Malaysia. The general approaches used for stuttering therapy in Malaysia are prolongation of speech, airflow management, regulation of breathing, gentle phonatory onset, shadowing, taping and speech rate reduction. The process of determining suitable therapy techniques may take months of repeated procedures that are costly and overly generalized. These techniques work to certain degree and the results depend very much on the particular case. Therefore, we implement a computer-based Malay stuttering therapy with the combination of different stuttering therapy techniques to help SLP in determining suitable techniques for each client. There are differences of

tense and lax vowels in pronunciation between Malay and English Language where Malay Language has only 6 vowel sounds, with no long-short distinction.

Our project implements DSP techniques to analyze speech signals and incorporates standard speech fluency shaping techniques that can be used as part of fluency rehabilitation regimen. The software runs under Windows XP on a computer equipped with multimedia capabilities. The software is developed using the Microsoft Visual C++ 6.0 Integrated Development Environment (IDE). The system software is developed as graphic user interface (GUI), which makes therapy user friendly. Our hypothesis is by doing comparison between client's and clinician's AMPs, the computer can help SLP to determine suitable technique. This will reduce cost and time for therapy. The example of AMPs is showed in Figure 1.

This software tool is similar to speech therapy tools only found in the speech pathology clinic in some countries and provides the added advantage of being available for home use by clients that own a personal computer. Three techniques implemented in our project are Shadowing [18, 19, 20], Metronome (Taping) and Delayed Auditory Feedback.

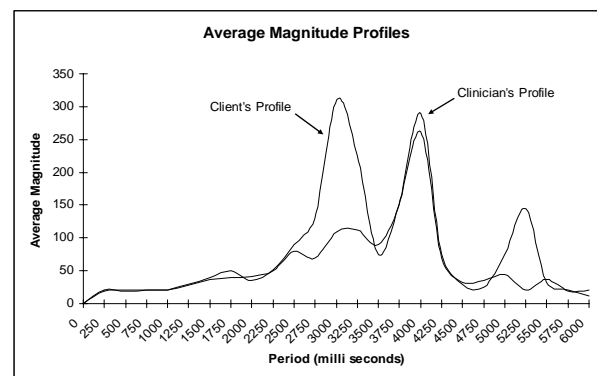


Figure 1. Examples of Average Magnitude Profiles of Client and Clinician

7. Reviews

The published literature on the clinical use and effectiveness of the computer-based devices described above consists of a few reports with small numbers of patients in uncontrolled case series. The results are somewhat mixed, but suggest a decrease in stuttering in some individuals performing reading tasks more so than monologue.

There is no one best treatment for stuttering for every client. Therefore, it is suggested that choosing different types of stuttering managements is a personal decision based on the unique values and treatment objectives of each individual person who stutters. Without a definitive treatment for developmental stuttering, the field of speech-language pathology continues to serve the stuttering population by offering as many valid and reliable treatments as possible; however, it is ultimately the client's responsibility to become informed of the various approaches towards stuttering management, and pick specific treatments

that best meets their needs. With this in light, the inclusion of a prosthetic computer-based device into a stuttering management program becomes a private choice that is determined by the personal values and treatment objectives of each individual client.

Normally, 2 to 3 months are required to determine suitable therapy technique for each client. These techniques work to certain degree and the results depend very much on the particular case. The treatment process may take months of repeated procedures that are costly and overly generalized. Therefore, implementation of a computer-based Malay stuttering therapy with combination of different stuttering therapy techniques is required to help SLP in determining suitable techniques for each client.

8. Conclusion

There are four classes of assistive and anti-stuttering computer-based devices available that have been employed in treating stuttering. Alter Auditory Feedback (AAF) and Speech Motor Production Patterns Alteration are the most commonly used principals in designing computer-based stuttering devices.

Our software is designed to help SLP in determining suitable techniques for clients in faster and more accurate manner. This will reduce cost and time for therapy session. We believe our software tool will improve the effectiveness and availability of stuttering therapy in Malaysia. We hope that our software tool will provide insights into the implementation of computer-based Malay stuttering therapy in Malaysia.

Our system is currently being tested clinically at Speech Pathology Centre of Hospital Sultanah Aminah. The effectiveness of the software is currently under evaluation. A report of the effectiveness of the software will be available after the tool has been thoroughly evaluated.

Acknowledgments

Thanks to Ministry of Science, Technology and Innovation of Malaysia for funding this project.

References

- [1] Charles Van Riper and Robert L. Erickson. 1996. *Speech Correction – An Introduction to Speech Pathology and Audiology*, 9th Edition. USA: A Simon & Schuster Company, Needham Heights.
- [2] Stuart A, Xia S, Jiang Y, Jiang T, Kalinowski J, Rastatter MP. 2003. Self-contained in-the-ear device to deliver altered auditory feedback: applications for stuttering. *Ann Biomed Eng* 31(2):233-7.
- [3] Van Borsel, J., Reunes, G., and Van den Bergh, N. 2003. Delayed auditory feedback in the treatment of stuttering: clients as consumers. *International Journal of Language and Communication Disorders* Vol. 38, No. 2, 119-129.
- [4] Richard M. Merson. 2003. *Auditory Sidetone and the Management of Stuttering: From Wollensak to SpeechEasy*, Research Report, Speech-Language Pathology Department of William Beaumont Hospital, Michigan, USA.
- [5] M. N. Hegde. 1998. *PocketGuide to Treatment in Speech-Language Pathology*, 4th Edition. UK: Singular.
- [6] Radford, N., Tanguma, J., Gonzalez, M., Nericcio, M.A., Newman, D. 2005. A Case Study of Mediated Learning, Delayed Auditory Feedback, and Motor Repatterning to Reduce Stuttering. *Perceptual and Motor Skills* 101: 63-71.
- [7] Kalinowski, J., Stuart, A., Rastatter, M., Miller, R., Zimmerman, S., & Shine, R. 1998. Examination of altered auditory feedback: Therapeutic implications. In *Proceedings of Second World Congress on Fluency Disorders*, 54-57. Nijmegen, The Netherlands: University Press Nijmegen.
- [8] MacLeod, J., Kalinowski, J., Stuart, A., & Armson, J. 1995. Effect of single and combined altered auditory feedback on stuttering frequency at two speech rates. *Journal of Communication Disorders* 28: 217-228.
- [9] Hyde, L. 2003. Comparison of the SpeechEasy and Casa Futura/Jabra fluency devices. In *Proceedings of the Canadian Association of Persons who Stutter*.
- [10] Grosser, J., Natke, U., Langefeld, S., & Kalveram, K. Th. 2001. Reduction in stuttering by delayed and frequency shifted auditory feedback: Effects of adaptation and sex differences. In *Proceedings of the Third World Congress of Fluency Disorders*, 422-426. Nyborg, Denmark. Nijmegen: Nijmegen University Press.
- [11] Joseph Kalinowski. SpeechEasy, Janus Development Corp. Greenville, North Carolina. 26 May 2006. Citing Internet Sources URL www.speecheasy.com.
- [12] Daniel Boone. Kay Elemetrics Corp. Lincoln Park, New Jersey, USA. 26 May 2006. Citing Internet Sources URL www.kayelemetrics.com.
- [13] Shames, G. H., and Wiig, E. H. 1990. *Human Communication Disorders, 3rd Edition*. Columbus: Merrill.
- [14] Ronald L. Webster. The Hollins Fluency System: FluencyNet. Hollins Communications Research Institute. 26 May 2006. Citing Internet Sources URL www.stuttering.org.
- [15] Marek Roland-Mieszkowski. 1994. Digital Recordings and Andrzej Czyzewski, and Bozena Kostek. In *Proceedings of the First World Congress on Fluency Disorders*. Germany.
- [16] National Association for Speech Fluency. Fluency Master. 26 May 2006. Citing Internet Sources URL www.stutteringcontrol.com.

- [17] Martin, R.R, Johnson, L. J., Siegel, G.M, Haroldson and S. K. 1985. Auditory Stimulation, Rhythm, and Stuttering. *Journal of Speech and Hearing Research* 28:497-498.
- [18] Kalinowski, J., Saltuklaroglu, T., Guntupalli, V, & Stuart, A. 2004. Gestural Recovery and the Role of Forward and Reversed Syllabic Repetitions as Stuttering Inhibitors in Adults. *Neuroscience Letter* 363: 144-149.
- [19] Adamczyk, B. 1994. Stuttering Therapy with the "Echo Method". *Journal of Fluency Disorders* 19:47.
- [20] Andrews, G. and Craig, A. 1982. Stuttering: Overt and Covert Measurement of the Speech of Treated Stutterers. *Journal of Speech and Hearing Disorders* 47:96-99.

The Accuracy and Efficiency Issues of Decouple Approach for Harmonic Power Flow Calculation

Agus Ulinuha^{1*}, **Mohammad A. S. Masoum**², **Syed M. Islam**³

Department of Electrical and Computer Engineering
Curtin University of Technology, GPO Box U1987 Perth 684, Western Australia, Fax: +61-8-9266 2584

¹ Tel: +61-8-9266 3034, E-mail: agus.ulnuha@postgrad.curtin.edu.au

² Tel: +61-8-9266 7901, E-mail: m.masoum@curtin.edu.au

³ Tel: +61-8-9266 2566, E-mail: s.islam@ece.curtin.edu.au

Abstract

Power flow calculation is the backbone of power system analysis and design. This is normally carried out by only considering the fundamental frequency. Due to extensive use of nonlinear loads that generate and inject harmonic into power system, harmonic frequencies are present and need to be considered. Unfortunately, unavoidable complexity and heavy computational burden are often encountered by involving nonlinear loads into the already complicated power flow calculation. This paper implements a decouple approach to overcome the problem. The couplings between harmonics are rationally disregarded and as a result, the calculations are separately performed for every harmonic order. This will greatly reduce the level of complexity and computation charge. However, the accuracy level of this technique is somehow questioned due mainly to the neglected harmonic couplings. To investigate the accuracy of the implemented decoupled harmonic power flow (DHPF) algorithm, simulation results are compared with those generated by standard packages (e.g., HARMFLOW and ETAP). The distorted IEEE 18-bus system is used for simulation purposes, while the nonlinear load involved in the system is modeled by harmonic current sources. The comparison is carried out in terms of accuracy and efficiency. It is shown that the decoupled approach offers the compromise between result accuracy and computation complexity.

Keywords:

Decouple, Harmonic Power Flow, Nonlinear Load and Power Quality

1. Introduction

Power flow calculation is the backbone of power system analysis and design. It generates the results that are normally required for further calculation of analysis and design. This is initially performed by formulating the network equation. Node-voltage method, which is the most suitable form for many power system analyses, is commonly used. The calculation can then be carried out by solving the following equation.

$$I_{bus} = Y_{bus} V_{bus} \quad (1)$$

Where I_{bus} and V_{bus} are vectors of bus injection current and bus voltage, respectively, while Y_{bus} is bus admittance matrix. Mathematically, power flow problem requires solution of simultaneous nonlinear equations and normally employs an iterative method, such as Gauss-Seidel and Newton-Raphson.

The aforementioned calculation is typically carried out by considering fundamental frequency. The extensive and ever increasing applications of nonlinear loads such as power electronic devices result in existence of higher components other than that of fundamental frequency, called harmonics. The nonlinear voltage-current relationship of these devices results in harmonic currents that propagate through the system and produce potentially dangerous harmonic voltages. This phenomenon has become a major concern for power quality and therefore harmonics must be included in the calculations to predict their effects and to avoid possible severe damages. However, taking harmonics into account will lead the calculations to be very complicated.

This paper presents a decouple approach for harmonic power flow calculation. Harmonics are included in the calculations with a reasonable computation cost. To show the accuracy of the implemented decoupled harmonic power flow (DHPF) algorithm, simulation results for the distorted IEEE 18-bus distribution system are compared with those

* Corresponding Author. E-mail: agus.ulnuha@postgrad.curtin.edu.au,
Tel: +61-8-9266 3034, Fax: +61-8-9266 2584

generated by HARMFLOW and ETAP. It is shown that the decoupled approach offers the compromise between result accuracy and computation complexity.

2. Harmonic Power Flow

Harmonic power flow was initially introduced by Xia and Heydt [1] by involving nonlinear loads in power flow calculation. Conventionally, power flow is formulated on the basis that power sources are system generators and power “sinks” are loads. Harmonic power flow, on the other hand, is more general in that loads may be the “source” of harmonic energy [2]. The ultimate source is system generators, but harmonic distortion that occurs at bus containing nonlinear load may be considered as a source of harmonic signal. In addition to some results normally generated by power flow, harmonic power flow also generates other results that can be used to quantify voltage distortion and to determine whether dangerous resonant problem exists.

The nature of the harmonics (e.g., orders, magnitudes and phases) strongly depends on the nonlinear load involved. Therefore, nonlinear load modeling has become an essential part of harmonic power flow calculation. Nonlinear loads can be modeled in time and/or frequency domain [3]. Time domain modeling is based on transient-state analysis while frequency domain modeling uses frequency-scan process to calculate the frequency response of a system. Time domain modeling requires detailed representation of the device that increases the problem complexity resulting in prohibitively long computation time. Therefore, frequency domain methods are commonly used for harmonic analysis to reduce the computation time. For nonlinear loads that can be presented as voltage-independent current sources, frequency domain model can be applied for harmonic power flow analysis [4]. Harmonic power flow calculations can generally be classified into coupled and decoupled methods. Couple approaches solve all of the harmonic orders simultaneously. This approach has good accuracy but leads to a greater computational cost as the problem becomes quite complicated. It also requires exact formulation of nonlinear loads that is sometimes practically unavailable resulting in limited applications [5]. On the other hand, decouple approach assumes that the coupling between harmonic orders can be rationally disregarded and, as a result, the calculation can be separately carried out for every harmonic order. Therefore, this approach requires less computational charge. In addition, since nonlinear loads are modeled with harmonic current or voltage sources, it is very easy to include them in the calculations using measured non-sinusoidal current and/or voltage waveforms. Decoupled approaches are not as accurate as the coupled techniques; however, they offer a compromise between computational complexity and result accuracy.

2.1. Decouple Approach for Harmonic Power Flow

At the fundamental frequency, system is modeled using the conventional approach where the admittance of line section between bus i and bus $i+1$ is expressed as follows.

$$y_{i,i+1} = \frac{1}{R_{i,i+1} + jX_{i,i+1}} \quad (2)$$

Where $R_{i,i+1}$ and $X_{i,i+1}$ are the respective resistance and inductance of line section between buses i and $i+1$. The magnitude and phase angle of bus voltage is then calculated using the following mismatch equations [6-8].

$$P_i \sum_{j=i}^{i+1} |Y_{ji}^l| |V_j^l| |V_i^l| \cos(\delta_i^l - \delta_j^l - \theta_{ji}^l) = 0 \quad (3)$$

$$Q_i \sum_{j=i}^{i+1} |Y_{ji}^l| |V_j^l| |V_i^l| \sin(\delta_i^l - \delta_j^l - \theta_{ji}^l) = 0 \quad (4)$$

Where

$$Y_{ji}^l = |Y_{ji}^l| \angle \theta_{ji}^l = \begin{cases} y_{ji}^l, & \text{if } j \neq i \\ y_{i-1,i}^l + y_{i+1,i}^l + y_{ci}^l, & \text{if } j = i \end{cases} \quad (5)$$

V_i^l and y_{ci}^l are the respective fundamental voltage and admittance of shunt capacitor at bus i , while P_i and Q_i are the respective total (linear and nonlinear) active and reactive powers at bus i . The power loss in the line section between buses i and $i+1$ may then be calculated by the following equation.

$$P_{loss(i,i+1)}^l = R_{i,i+1} \left(|V_{i+1}^l| |V_i^l| |y_{i,i+1}^l| \right)^2 \quad (6)$$

At harmonic frequencies, power system is modeled as combination of passive elements and current sources [6]. The generalized model is suggested for a linear load, which is composed by a resistance in parallel with an inductance to account for the respective active and reactive loads at fundamental frequency. Nonlinear loads, in general, are considered as ideal harmonic current sources that generate harmonic currents and inject them into the system [9]. The admittance-matrix-based harmonic power flow is the most widely used method as it is based on the frequency-scan process [10]. In this approach, admittance of system components will vary with the harmonic order. If skin effect is ignored at higher frequencies, the resulting n^{th} harmonic frequency load admittance, shunt capacitor admittance and feeder admittance are respectively given by the following equations [6-8, 11-14].

$$y_{li}^n = \frac{P_{li}}{|V_i^l|^2} - j \frac{Q_{li}}{n|V_i^l|^2} \quad (7)$$

$$y_{ci}^n = n y_{ci}^n \quad (8)$$

$$y_{i,i+1}^n = \frac{1}{R_{i,i+1} + jnX_{i,i+1}} \quad (9)$$

Where P_{li} and Q_{li} are the respective active and reactive

linear loads at bus i . The nonlinear load is treated as harmonic current sources and the n^{th} harmonic current injected at bus i introduced by the nonlinear load is derived as follows:

$$I_i^1 = [(P_{ni} + jQ_{ni}) / V_i^1]^* \quad (10)$$

$$I_i^n = C(n)I_i^1 \quad (11)$$

Where I_i^1 is the fundamental current and I_i^n is the n^{th} harmonic current determined by $C(n)$, which is the ratio of the n^{th} harmonic to the fundamental current. $C(n)$ can be obtained by field test and Fourier analysis for all customers along the distribution feeder [6, 8, 14].

For decouple harmonic power flow calculation, loop equations are written at each harmonic frequency of interest. Each loop is formed including the source nodes. After modifying admittance matrix and the associated harmonic current, the harmonic load flow problem can then be calculated by the following equation [10, 13, 14].

$$Y^n V^n = I^n \quad (12)$$

At any bus i , the rms voltage is defined by:

$$|V_i| = \sum_{n=1}^N |V_i^n|^2 \quad (13)$$

Where N is the maximum harmonic orders considered. After solving the load flow for different harmonic orders, the distortion of voltage indicated by the total harmonic distortion at bus i (THD_{vi}) is expressed by the following equation.

$$THD_{vi}(\%) = \frac{\sum_{n \neq 1}^N |V_i^n|^2}{|V_i^1|} \times 100\% \quad (14)$$

At the n^{th} harmonic frequency, real power loss in the line section between buses i and $i+1$ is expressed below [7, 8, 14].

$$P_{loss(i,i+1)}^n = R_{i,i+1} (|V_{i,i+1}^n - V_i^n| |y_{i,i+1}^n|)^2 \quad (15)$$

The total power loss of the system for all harmonics is therefore given by the following equation.

$$P_{loss}^n = \sum_{n=1}^N \sum_{i=0}^{m-1} P_{loss(i,i+1)}^n \quad (16)$$

Where m is the total number of buses. The computation procedure of the proposed approach is given by Figure 1.

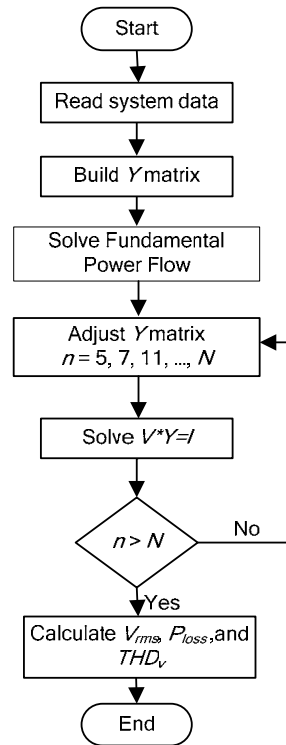


Figure 1. Decouple Harmonic Power Flow Calculation

3. Result and discussion

The distorted IEEE 18-bus distribution system [15] as shown in Figure 2 is simulated using the proposed decoupled harmonic power flow (DHPF). This system includes a 3 MW 6-pulse converter as the nonlinear load, which is modeled by current sources. The non-sinusoidal current waveform injected by this nonlinear load is shown in Figure 3 and its harmonic contents are presented in the Appendix. The proposed approach is coded using MATLAB version 7.0.1 R14 and is run in a desktop PC with Pentium 4 Intel 3.0 GHz processor and 512 MB RAM.

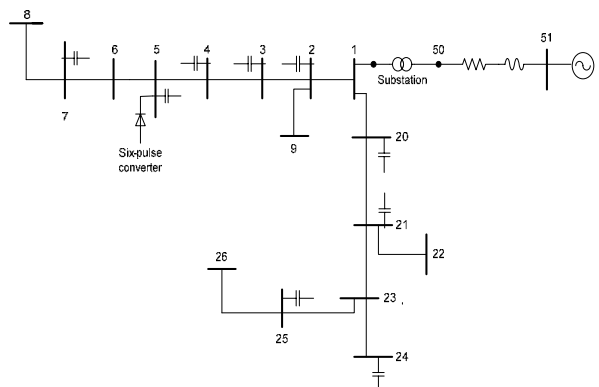


Figure 2. The simulated IEEE 18-bus distorted distribution system

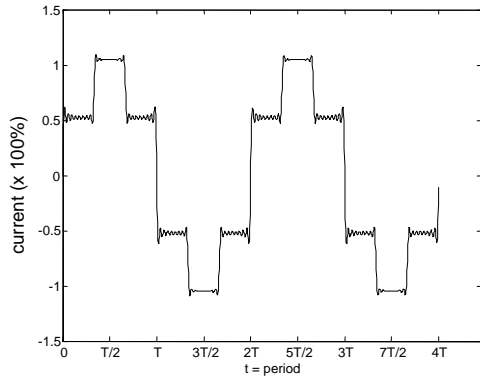


Figure 3. The non-sinusoidal current waveform of nonlinear load used in this paper

The generated results of the proposed DHPF including fundamental voltage (V_{fund}), rms voltage (V_{rms}), and THD of voltage (THDv) are shown in Table 1. For verification of accuracy, these results are compared with those generated by two standard software packages (HARMFLOW and ETAP). The main reason of results comparison is to demonstrate their accuracy.

Table 1. Simulation results of the Harmonic Power Flow

Bus	Vfund (p.u.)	Vrms (p.u.)	THDv (%)
1	1.0545	1.0550	2.79
2	1.0511	1.0516	3.13
3	1.0456	1.0462	3.49
4	1.0425	1.0432	3.67
5	1.0359	1.0368	4.23
6	1.0348	1.0358	4.31
7	1.0326	1.0336	4.52
8	1.0268	1.0278	4.52
9	1.0496	1.0501	3.13
20	1.0505	1.0508	2.69
21	1.0496	1.0502	3.44
22	1.0479	1.0485	3.44
23	1.0451	1.0465	5.22
24	1.0485	1.0506	6.30
25	1.0419	1.0437	5.93
26	1.0415	1.0433	5.93
50	1.0501	1.0501	0.26
51	1.0500	1.0504	0.26

3.1. Comparison with HARMFLOW

The deviations of results generated by the proposed DHPF from those generated by HARMFLOW are indicated in Table 2. Maximum and average deviations are also provided in Table 3. In addition, figures 4 and 5 illustrate the deviations of V_{rms} and THDv, respectively. Tables 2-3 and Figures 4-5 indicate that, in general, the results generated by the proposed approach are fairly close to the results generated by HARMFLOW. As expected, there are some slight differences at some buses due to the neglected harmonic coupling by the proposed DHPF. As the THD

values are calculated in percentages of the fundamental voltage, their deviations values are relatively large.

Table 2. Deviation of results from those generated by HARMFLOW

Bus	Deviation (%) of		
	V_{fund}	V_{rms}	THDv
1	0.0000	0.0095	0.33308
2	0.0095	0.0000	0.05890
3	0.0096	0.0191	0.60019
4	0.0096	0.0288	0.83510
5	0.0097	0.0675	1.28137
6	0.0097	0.0676	1.31939
7	0.0097	0.0774	1.41657
8	0.0097	0.0778	1.42473
9	0.0000	0.0000	0.05879
20	0.0095	0.0000	0.00808
21	0.0095	0.0095	0.28274
22	0.0000	0.0095	0.28347
23	0.0000	0.0286	0.49441
24	0.0000	0.0380	0.49794
25	0.0096	0.0287	0.51851
26	0.0096	0.0383	0.51908
50	0.0000	0.0000	0.12646
51	0.0000	0.0000	0.24472

Table 3. The Maximum and Average Deviations of Table 2

	V_{fund}	V_{rms}	THDv
Maximum Deviation (%)	0.0097	0.0778	1.42473
Average Deviation (%)	0.0059	0.0278	0.57242

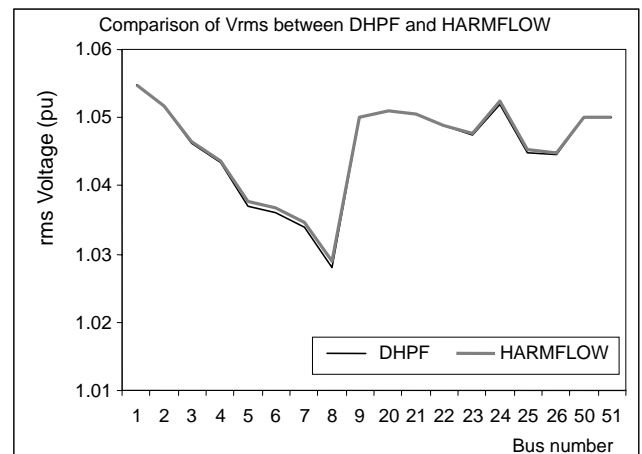


Figure 4. Comparison of V_{rms} values computed by DHPF and HARMFLOW

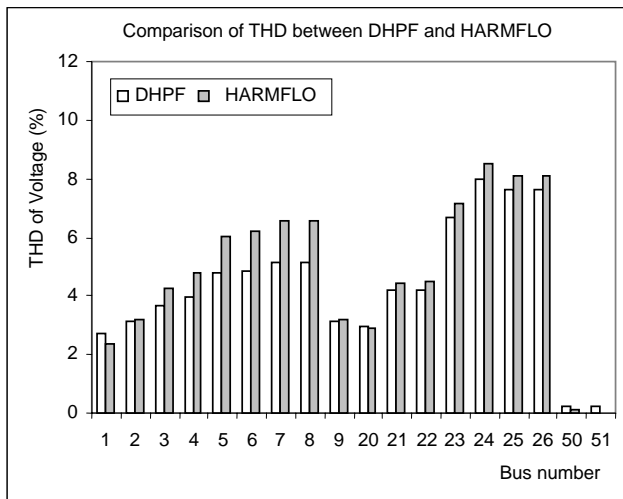


Figure 5. Comparison of THD values computed by DHPF and HARMFLOW

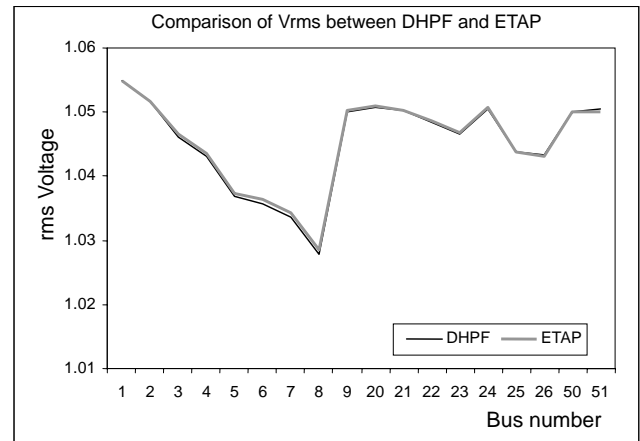


Figure 6. Comparison of Vrms values computed by DHPF and and ETAP

3.2. Comparison with ETAP

Comparison of the simulation results with those generated by ETAP is shown in Table 4. The maximum and average deviations are summarized in Table 5, indicating fine agreements between the results and justifying the accuracy of the proposed approach for harmonic power flow calculations. Figures 6 and 7 illustrate these verifications.

Table 4. Deviation of results from those generated by ETAP

Bus	Deviation (%) in		
	Vfund	Vrms	THDv
1	0.0000	0.0095	0.12390
2	0.0000	0.0095	0.18930
3	0.0000	0.0287	0.66160
4	0.0000	0.0383	0.87260
5	0.0000	0.0579	1.30290
6	0.0000	0.0579	1.32050
7	0.0000	0.0774	1.37540
8	0.0000	0.0681	1.37600
9	0.0000	0.0095	0.18930
20	0.0000	0.0095	0.09630
21	0.0000	0.0095	0.24390
22	0.0000	0.0095	0.24350
23	0.0000	0.0191	0.26580
24	0.0000	0.0095	0.18060
25	0.0000	0.0096	0.21690
26	0.0384	0.0192	0.21670
50	0.0000	0.0000	0.04480
51	0.0000	0.0381	0.07520

Table 5. The Maximum and Average Deviations of Table 4

	Vfund	Vrms	THDv
Maximum Deviation (%)	0.0384	0.0774	1.3760
Average Deviation (%)	0.0021	0.0267	0.4997

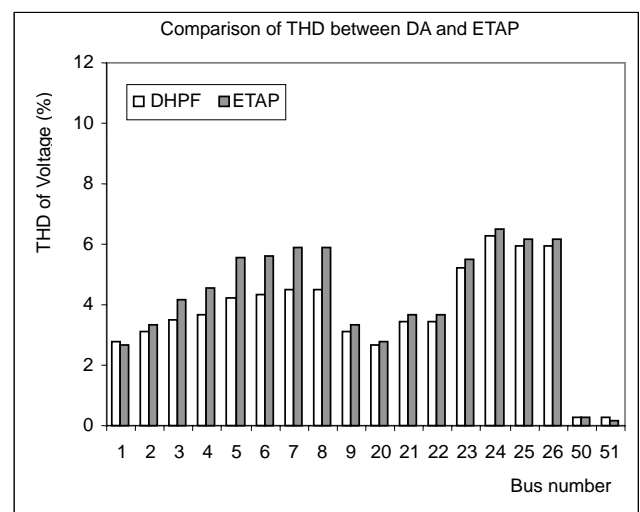


Figure 7. Comparison of THD values computed by DHPF and ETAP

The aforementioned comparisons confirm that the results generated by the decouple approach is fairly accurate. The proposed DHPF has also been employed to simulate a distorted IEEE 9-bus system [8, 9] and fairly accurate results were generated.

3.3. Efficiency of the proposed DHPF

In the proposed DHPF algorithm, calculations are separately performed for every harmonic order. Therefore, this approach is very simple compared with the coupled harmonic power flow algorithm and can be used to simulate large distorted distribution systems without any convergence difficulties. The algorithm was also used to simulate a 300-bus distorted system. Most harmonic power flow algorithms are not capable of simulating large systems (e.g., with hundreds of buses). DHPF is therefore suitable for simulating large distorted distribution systems

The nonlinear load modeling is another advantage of the decouple approach. This approach simply estimates nonlinear loads as harmonic current sources that can be easily obtained from measurements. In contrast, the coupled approaches require exact models for nonlinear loads that are not usually available.

Computation time is another aspect that needs to be considered. DHPF has less computational burden and will require less computing time compared with techniques that consider harmonic coupling.

4. Conclusion

The application of decouple approach for harmonic power flow is presented. Comparisons of the generated results with those computed by standard software packages are presented and discussed. Main conclusions are:

- The formulation and nonlinear load modeling of the decouple approach leads the calculation to be simple.
- From the comparisons, the decouple approach offers a compromise between the result accuracy and calculation complexity.
- The decouple approach can be extensively applied due to its simple nonlinear load modeling and is more practical as it can handle large systems.

Appendix: Harmonic Model of 6-pulse Converter

Nonlinear loads are modeled as decoupled harmonic current sources. Table A gives the current magnitudes (as percentages of the fundamental current) used to model 6-pulse converter loads. Harmonic phase angles are assumed to be zero.

Table A. Magnitude (%) of harmonic currents for 6-pulse converter

Order	Mag.	Order	Mag.	Order	Mag.
1	100	19	5.3	37	2.7
5	20	23	4.3	41	2.4
7	14.3	25	4	43	2.3
11	9.1	29	3.4	47	2.1
13	7.7	31	3.2	49	2
17	5.9	35	2.8		

Acknowledgment

The first author would like to gratefully acknowledge that this doctoral research is sponsored by SPMU-TPSDP Universitas Muhammadiyah Surakarta under contract number: 022/SD-TK/SPMU-UMS/VI/04.

References

[1] Xia, D., and Heydt, G. T. 1982. Harmonic power flow studies, part I - Formulation and Solution. *IEEE Trans. on Power Apparatus and System*. 101: 1257-1265.

[2] Semlyen, A., and Shlash, M. 2000. Principles of modular harmonic power flow methodology. *IEE Proceedings Generation, Transmission and Distribution*. 147: 1-6.

[3] Moreno Lopez de Saa, M. A., and Usaola Garcia, J. 2003. Three-phase harmonic load flow in frequency

and time domains. *IEE Proceedings - Electric Power Applications*, 150: 295-300.

- [4] Hong, Y. Y.; Lin, J. S., and Liu, C. H. 2000. Fuzzy harmonic power flow analyses. In *International Conference on Power System Technology*.
- [5] Williams, S. M.; Brownfield, G. T.; and Duffus, J. W. 1993. Harmonic propagation on an electric distribution system: field measurements compared with computer simulation. *IEEE Transactions on Power Delivery*. 8: 547-552,.
- [6] Chin, H. C. 1995. Optimal shunt capacitor allocation by fuzzy dynamic programming. *Electric Power Systems Research*, 35: 133-139.
- [7] Baghzouz, Y., and Ertem, S. 1990. Shunt capacitor sizing for radial distribution feeders with distorted substation voltage. *IEEE Trans. on Power Delivery*. 5: 650-657.
- [8] Baghzouz, Y. 1991. Effects of nonlinear loads on optimal capacitor placement in radial feeders. *IEEE Trans. on Power Delivery*. 6: 245-251.
- [9] Yu, X.-m.; Xiong, X.-y.; and Wu, Y.-w. 2004. A PSO-based approach to optimal capacitor placement with harmonic distortion consideration. *Electric Power Systems Research*. 71: 27-33.
- [10] Teng, J.-H., and Chang, C.-Y. 2003. Fast harmonic analysis method for unbalanced distribution systems. In *IEEE Power Engineering Society General Meeting*.
- [11] Masoum, M. A. S.; Jafarian, A.; Ladjevardi, M.; Fuchs, E. F.; and Grady, W. M. 2004. Fuzzy approach for optimal placement and sizing of capacitor banks in the presence of harmonics. *IEEE Transactions on Power Delivery*, 19: 822 - 829.
- [12] Masoum, M. A. S.; Ladjevardi, M.; Jafarian, A.; and Fuchs, E. F. 2004. Optimal Placement, Replacement and Sizing of Capacitor Banks in Distorted Distribution Networks by Genetic Algorithms. *IEEE Transactions on Power Delivery*, 19: 1794-1801.
- [13] Ghose, T.; and Goswami, S. K. 2003. Effects of unbalances and harmonics on optimal capacitor placement in distribution system. *Electric Power Systems Research*. 68: 167-173.
- [14] Chung, T. S.; and Leung, H. C. 1999. A genetic algorithm approach in optimal capacitor selection with harmonic distortion considerations. *International Journal of Electrical Power & Energy Systems*. 21: 561-569.
- [15] Grady, W. M.; Samotyj, M. J.; and Noyola, A. H. 1992. The application of network objective functions for actively minimizing the impact of voltage harmonics in power systems. *IEEE Transactions on Power Delivery*. 7: 1379-1386.

Design of Multimode Interference Optical Splitter Based on BenzoCyclobutene (BCB 4024-40) Polymer

Mohd Haniff Ibrahim^{*}, Norazan Mohd Kassim^{*}, Abu Bakar Mohammad^{*} & Mee-Koy Chin[#]

^{*}Photonic Research Group,
Faculty of Electrical Engineering,
Universiti Teknologi Malaysia.
81310 Skudai, Johor Malaysia.
haniff@fke.utm.my

[#]Photonic Research Centre,
School of Electrical and Electronic Engineering,
Nanyang Technological University. Singapore

Abstract

1x2 and 1x3 planar optical splitter based on BenzoCyclobutene (BCB 4024-40) polymeric material are proposed. The design is based on symmetric interference mechanism and simulated using the combined technique of effective index method and commercial BPM-CAD software from Optiwave™. A ridge waveguide of BCB 4024-40 on BK7 glass substrate is employed as the simulated structure. The simulation at 1550 nm optical wavelength shows an insertion loss of 2.75 dB and 4.73 dB for 1x2 and 1x3 splitter respectively. The uniformity is shown to be less than 0.5 dB. This provides useful idea on the applicability of BCB 4024-40 to be realized as an optical splitter.

Keywords:

BenzoCyclobutene polymer, optical splitter, chemical etching technique, multimode interference (MMI), polymer optical waveguide.

I. Introduction

Optical splitters are key components in photonic integrated circuits both for signal routing and signal processing. Standard splitters that realized based on X- and Y- junctions design suffer from high reflection and radiation loss due to branching complexity [1]. In the past couple of years, there has been a growing interest in the application of the multimode interference (MMI) effect in integrated optics [2][3]. Due to their excellent properties and ease of fabrication, MMI based splitters are becoming increasingly popular for various applications in integrated optics. Their main advantages are ultra-compact size, low loss, and large fabrication tolerances [2]. They are quite easy to design and are compatible with both weakly-guided and strongly-guided structures [3].

The demand in optical networking for photonic components that meet performance criteria as well as economic requirements has opened the door for technologies capable of high yield and low cost manufacturing while delivering high performance and enabling unique functions. Polymeric materials are

particularly attractive because of their ability to be processed rapidly, cost-effectively, and with high yields. Of that matter, polymer material has been accepted as a new generation material for optical integrated circuit due to its various advantages as compared to other optical materials.

In this paper, we will report on the design of 1x2 and 1x3 optical splitters using organic BenzoCyclobutene (BCB 4024-40) polymer from The Dow™ Chemical Company. The motivation of this paper is to visualize the applicability of chemical etching based polymer, BCB 4024-40 as an optical material as it may lead to extremely lower cost manufacturing of optical devices.

2. Theory

MMI splitters work on the principle of self imaging effect, a property of multimode waveguides by which an input field is reproduced in single or multiple images at periodic intervals along the propagation direction of the guide [2]. As shown in figure 1, the typical structure of MMI consists of (i) single mode input waveguides, (ii) a

multimode section and (iii) single mode output waveguides of the input geometry.

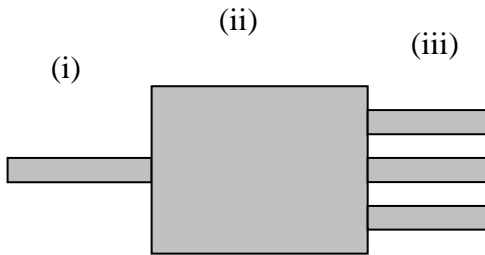


Figure 1: Schematic of MMI optical splitter.

The single mode optical field excites a weighted sum of the lateral modes of the multimode region, in which the relative weights of the modes being determined by the input modal field shape. The lateral modes have different propagation constants. At certain distances, a beating phenomenon occur where constructive interference between the modes will produces single or multiple self images of the input field. These images can then be coupled to single mode waveguides at the end of multimode region which can be further used as a passive N way optical splitter. In a special case whereby, the odd modes are not excited in the multimode waveguide, the 1 to N beam splitters can be realized with multimode waveguides four times shorter [3]. This condition can be achieved by centre- feeding the multimode waveguide with a symmetric field profile [2]. The imaging is obtained by linear combinations of the even modes (symmetrical) and this type of MMI mechanism is called symmetric interference.

Light intensities in the MMI waveguide exhibit various distributions as they propagate at the positions relative to the beat length, L_π defined by [2]

$$L_\pi = \frac{\pi}{\beta_0 - \beta_1} \frac{4nW_e}{3\lambda_0} \quad (1)$$

where β_0 and β_1 are propagation constants of the fundamental and the first order lateral modes respectively, λ_0 is a free space wavelength, n is an effective index and W_e is an effective width of the multimode waveguide. According to the symmetric interference mechanism, the single images of the input field will be obtained at the multimode section length of [2]

$$L = p \frac{3L_\pi}{4} \quad (2)$$

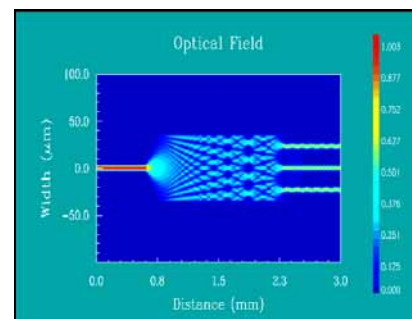
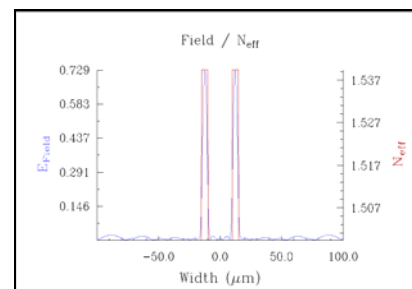
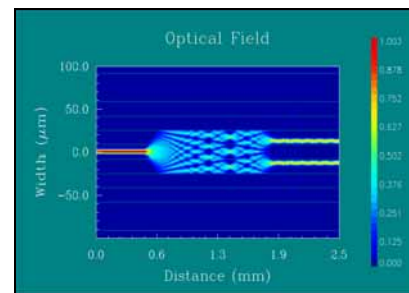
for $p=1,2,\dots$
while the N fold images are obtained at distances [2]

$$L = \frac{p}{N} \frac{3L_\pi}{4} \quad (3)$$

3. Simulation Results

In our design, we have considered a MMI splitter based on a ridge structure of BCB 4024-40 polymer on BK7 glass as a substrate and thin layer of SiO₂ as a cover. An integration of effective index method [7] and two-dimensional beam propagation method (2D-BPM) from Optiwave® is employed for the analysis. The design analysis is based on our recent material characterization work [6]. It was shown that for optical window of 1550 nm, the refractive index of the polymer is 1.5556 and to produce a single mode structure, a ridge thickness of 4 μm and mask opening of 4 μm are adopted.

Due to our analysis, for 50 μm wide multimode section, the 3 dB length for 1x2 splitter is 1321 μm , while for 1x3 splitter, the required multimode length to produce triple outputs is 1761 μm for 70 μm wide. The fields at the output of these splitters using the 2D-BPM analysis are shown in figure 2. It is seen that the insertion loss for 1x2 splitter is 2.75 dB, while for 1x3 splitter, the design works well with insertion loss of 4.57 dB.



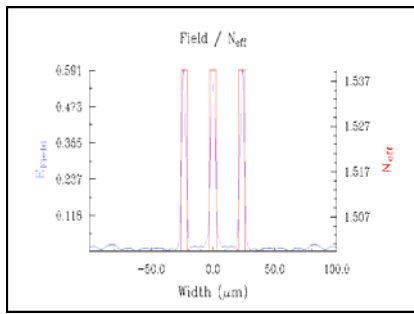


Figure 2: 2D-BPM analysis of 1x2 and 1x3 optical splitters.

4. Discussion and Conclusion

A performance of optical splitter is classified by imbalance parameter, which is defined as a measure of deviation of the splitting ratio from the ideal case [3]. It can be defined as:

$$\text{Imbalance} = 10 \log_{10} \frac{P_i}{P_{ref}} \quad (4)$$

where P_i is the optical power in a given output waveguide and P_{ref} is the optical power of an arbitrary chosen reference port.. Our simulation results show that the simulated imbalance of 1x2 splitter is always less than 0.2 dB, while for 1x3 splitter, the average measured imbalance is always better than 0.5 dB.

The obtained results shows the applicability of chemical etching based polymer, BCB 4024-40 to be employed as an optical device material, whilst reducing the realization cost.

ACKNOWLEDGMENT

M.H. Ibrahim, N.M. Kassim and A.B. Mohammad wish to thank Ministry of Science, Technology and Innovation of Malaysia (MOSTI) for funding this research work.

References

1. H.P. Chan, S.Y. Cheng and P.S. Chung. Low loss wide angle symmetric Y-branch waveguide, *Electronics Letters*, 32(7), 1996, pp. 652-654.
2. L.B. Soldano and E.C.M. Pennings. Optical multimode interference devices based on self imaging: principles and applications, *Journal of Lightwave Technology*, 13(4), 1995, pp.615-627.
3. L.B. Soldano, F.B. Veerman, M.K. Smit, B.H. Verbeck, A.H. Dubost and E.C.M. Pennings. Planar monomode optical couplers based on multimode interference effects, *Journal of Lightwave Technology*, 10(12), 1992, pp.1843-1850.
4. J F. Wang, J. Yang, L. Chen, X. Jiang and M. Wang. Optical switch based on multimode interference coupler, *IEEE Photonics Technology Letters*, 18(2), 2006, pp. 421-423.
5. R.M. Jenkins, R.W.J. Deveraux and J.M. Heaton. Waveguide beam splitters and recombiners based on multimode propagation phenomena, *Optics Letters*, 17(14), 1992, pp. 991-993.
6. Mohd Haniff Ibrahim, Norazan Mohd Kassim, Abu Bakar Mohammad and Mee-Koy Chin. Prism Coupling Measurement of BenzoCyclobutene (BCB 4024-40) Polymer for Photonic Devices Application, *Submitted and Accepted for Publication in Jurnal Elekrika*, June 2006 Issue.
7. Kassim et. al. Single Mode Rib Optical Waveguides Modelling Technique, *RF and Microwave Conference Proceedings*, October 2004, pp. 272-276.

Analysis and Classification of Myocardial Infarction Tissue from Echocardiography Images Based on Texture Analysis

Nazori Agani¹⁾, S.A.R. Abu-Bakar²⁾, and Sheikh Hussain Sheikh Salleh²⁾

¹⁾Department of Electrical Engineering Universitas Budi Luhur
Jl. Raya Ciledug, Jakarta Selatan, Indonesia (12260)
Telp. (021)5853753, Fax.(021)5853752
nazori@bl.ac.id

²⁾Department of Electrical Engineering Univeriti Teknologi Malaysia
81310 Skudai, Johor, Malaysia
syed@fke.utm.my, hussain@fke.utm.my

Abstract

Texture analysis is an important characteristic for automatic visual inspection for surface and object identification from medical images and other type of images. This paper presents an application of wavelet extension and Gray level co-occurrence matrix (GLCM) for diagnosis of myocardial infarction tissue from echocardiography images. Many of applications approach have provided good result in different fields of application, but could not implemented at all when texture samples are small dimensions caused by low quality of images. Wavelet extension procedure is used to determine the frequency bands carrying the most information about the texture by decomposition images into multiple frequency bands and to form an image approximation with higher resolution. Thus, wavelet extension procedure offers the ability to robust feature extraction in images. The gray level co-occurrence matrices are computed for each sub-band. The feature vector of testing image and other feature vector as normal image classified by Mahalanobis distance to decide whether the test image is infarction or not.

Keywords: wavelet extension, feature extraction, myocardial infarction, co-occurrence matrices.

1. Introduction

Textures provide important role for automatic visual inspection. The ability to represent is the single most important step in the development of system for measuring the similarity of textures and segmenting images on the basis of differences in textures. Their analysis is fundamental to many applications such as industrial monitoring of product quality control, remote sensing of earth resources, and medical diagnosis with computer tomography. Much research work has been done on texture analysis, such as classification, compression, retrieval and segmentation for last three decades. Despite the effort, texture analysis is still considered an interesting but difficult problem in image processing [4], [6],[11],[12].

Acute myocardial infarction is caused by the obstruction of a coronary artery by a thrombus, leading to irreversible damage of the heart muscle (myocardium). Echocardiography is a diagnostic test that uses ultrasound waves to create an image of the heart muscle. It may show such abnormalities as poorly functioning heart valves or damage to the heart tissue after acute myocardial

infarction. Texture characteristic of ultrasound image is low quality, caused by noise, low frequency and small dimension.

Wavelet extension algorithm is proposed for improving quality of texture result of new image with higher resolution. Aleksandra Mojsilovic *et.al* [9] showed that, from the texture characterization perspective, the proposed decomposition scheme performs more efficient energy distribution of an image, and the first-order, second-order, and higher-order statistics calculated on the expanded images can be used as reliable texture description for classification purpose. Gray level co-occurrence matrix, one of the most known texture analysis methods, estimates image properties related to second order statistics. Mari Patrio *et.al.*[14] have used the feature extracted from GLCM with the problem of how to guarantee even quality within a set of rock plates. A.L. Amer *et.al.*[10] proposed a method, namely, the sub-band domain co-occurrence matrix to solved the texture defect detection problem. Therefore, this research proposed a new combines concept of wavelet extension transform with GLCM texture feature were used for diagnosis of

myocardial infarction tissue and retrieval in small dimension images

The goal of this paper is to establish the algorithms for texture analysis, which can be detected by texture image as distinguishing a textural normal myocardium from textural infarcted.

This paper is organized as follow. Section 2 introduces background theory of wavelet and co-occurrence matrix. Experimental results are present in section 3, and finally, in section 4, includes the concluding remarks.

2. Methodology

The proposed defect detection and texture retrieval system consist of two stages [10]: (i) The feature extraction part which first utilizes the wavelet extension procedure to decompose textured image into sub-bands and GLCM procedure to computed energy, entropy, contrast and inverse difference moment for each sub-bands (ii) The detection part (texture classification) which is a mahalanobis distance classifier being trained by defect free samples (see fig.1).

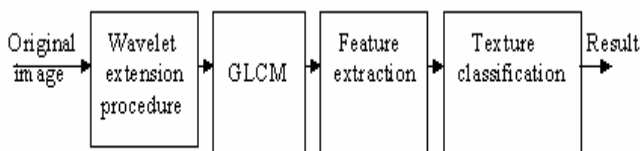


Figure 1: Block diagram

A. Review of Wavelet Transform

The wavelet transform is define as a decomposition of signal $f(t)$ with a family of real orthonormal bases $\psi_{m,n}(t)$ generated from a kernel function $\psi(t)$ by dilations and translations [5],[8],[13]:

$$\psi_{m,n}(t) = 2^{-m/2} \psi(2^{-m}t - n)$$

(1)

where j and k are integers.

The multiresolution formulation needs two closely related basic functions. In addition to the mother wavelet $\psi(t)$, we will need another basic function, called the scaling function $\phi(t)$. $\phi(t)$ can be expressed in term of weighted sum of shifted $\phi(2t - k)$ as [3]:

$$\phi(t) = \sqrt{2} \sum_k h(k) \phi(2t - k)$$

(2)

where $h(k)$'s are the scaling function (lowpass) coefficients and the mother wavelet $\psi(t)$ is related to the scaling function via

$$\psi(t) = \sqrt{2} \sum_n g(k) \phi(2t - k)$$

(3)

where $g(k)$'s are the wavelet (highpass) coefficients. They are required by orthogonality to be related to the scaling coefficients by

$$g(k) = (-1)^k h(1 - k)$$

(4)

The mother wavelet $\psi(t)$ is good at representing the detail and high-frequency parts of a signal. The scaling function $\phi(t)$ is good at representing the smooth and low-frequency parts of the signal.

The 1-D multiresolution wavelet decomposition can be easily extended to two dimensions by introducing separable 2-D scaling and wavelet functions as the tensor product of their one-dimensional complements.

$$\phi_{LL}(x, y) = \phi(x) \phi(y)$$

$$\psi_{LH}(x, y) = \phi(x) \psi(y)$$

$$\psi_{HL}(x, y) = \psi(x) \phi(y)$$

$$\psi_{HH}(x, y) = \psi(x) \psi(y)$$

The corresponding filter coefficient are

$$f_{LL}(x, y) = h(x) h(y)$$

$$f_{LH}(x, y) = h(x) g(y)$$

$$f_{HL}(x, y) = g(x) h(y)$$

$$f_{HH}(x, y) = g(x) g(y)$$

where the first and second subscripts denote, respectively, the lowpass and highpass filtering along the row and column direction of the image.

Figure 2 shows how to implement the wavelet decomposition of an image. After the decomposition, four subbands, LL, LH, HL and HH subbands, which represent the average, horizontal, vertical, and diagonal information respectively.

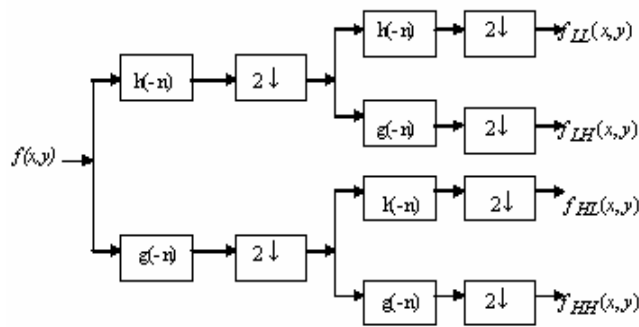


Figure 2: One stage in multiresolution image decomposition

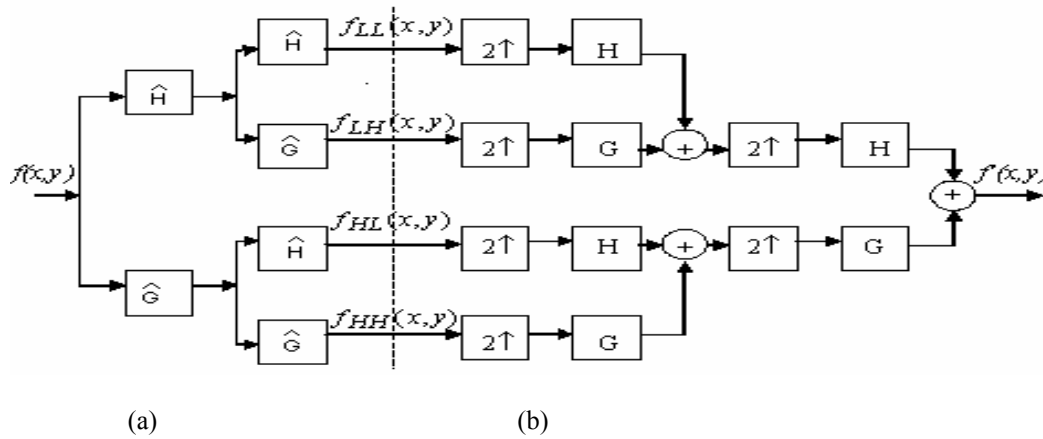


Figure 3: Block diagrams illustrating the complete wavelet decomposition-extension procedure (a) the composition part and (b) the extension (synthesis) algorithm.

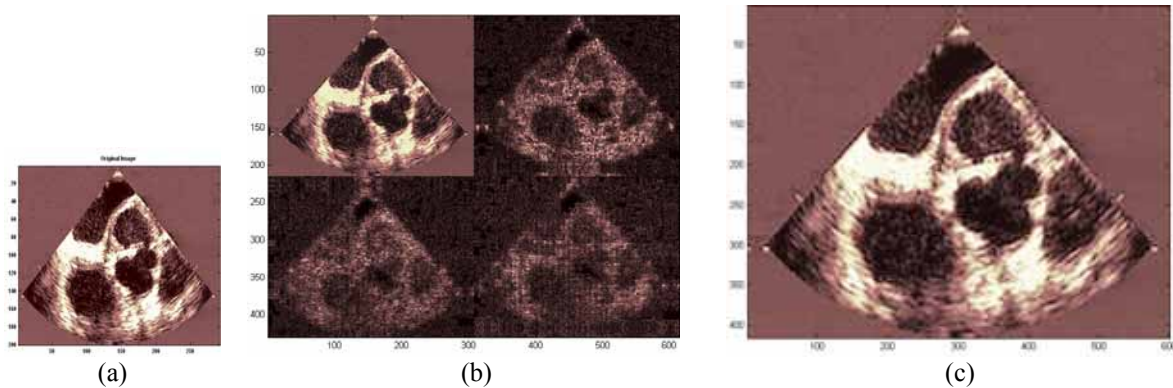


Figure 4: The result of the complete decomposition extension procedure for one representative ultrasound image of a human heart. (a). Original image (b) after the decomposition (c) Synthesized (reconstruction) image with two times higher resolution.

In order, we use a procedure called wavelet image extension. The application of the procedure is illustrated by the block diagram in Figure 3a. These four images are used as the input into the extension (interpolation) procedure, which is illustrated by the block diagram in Figure 3b.

B. Review of Co-occurrence Matrices

The co-occurrence matrix is defined by a distance and an angle, and its mathematical definition is

$$P_d[i, j] = |\{[r, c] : I[r, c] = i \text{ and } I[r + dr, c + dc] = j\}|$$

where d be a displacement vector (dr, dc) specifying the displacement between the pixel having values i and the

pixel having value j , dr is a displacement in rows (downward) and dc is a displacement in columns (to the right) and I denote an image of size $N \times N$ with G gray values [10].

Texture classification can be based on criteria (feature) derive from the occurrence matrices.

1). Entropy

$$ENT = \sum_i \sum_j p(i, j) \log p(i, j)$$

(5)

2). Contrast

$$CON = \sum_i \sum_j (i - j)^2 p(i, j)$$

(6)

3). Angular Second Moment

$$ASM = \sum_i \sum_j \{ p(i, j) \}^2$$

(7)

4). Inverse Difference Moment

$$IDM = \sum_i \sum_j \frac{1}{1 + (i - j)^2} p(i, j)$$

(8)

In Equation (5) – (8), $p(i, j)$ refers to the normalized entry of the co-occurrence matrices. That is

$$p(i, j) = \frac{P_d(i, j)}{R}$$

(9)



Figure 6: a. Normal myocardium zone (16x16), b. Infarcted myocardium zone (16x16).

where R is the total number of pixel pairs (i, j) . For a displacement vector $d = (dr, dc)$ and image of size $N \times M$ R is given by $(N-dr)(M-dc)$.

3. Experimental results

The experiments in this part are used texture image from ultrasound images taken from 15 patients, obtained from clinical hospital. For each patient to be analyzed, five tissue samples are taken from ultrasound image segments corresponding to area not affected (2) by infarction, and five tissue samples taken from image segments corresponding to the infarcted (1) area of myocardium.

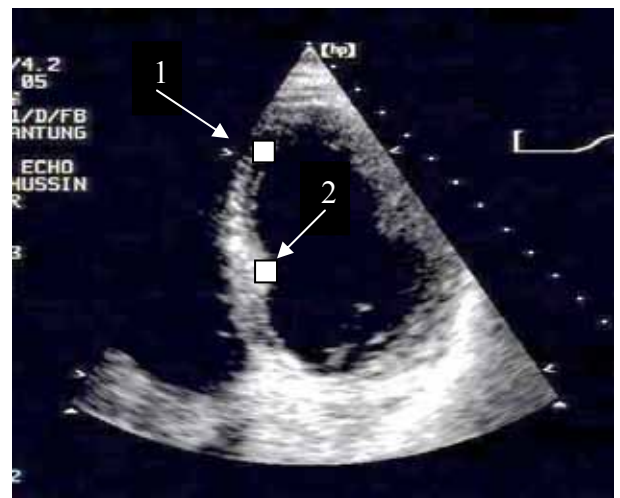
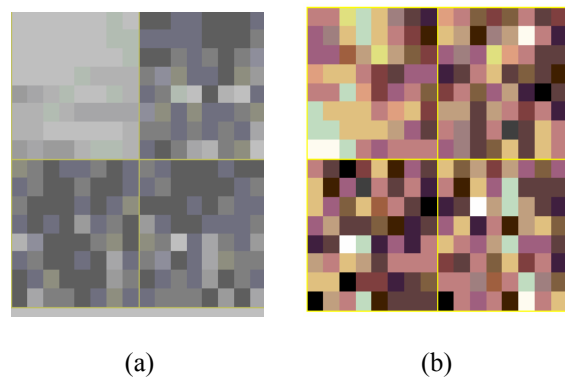
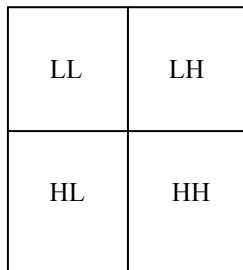


Figure 5: A typical ultrasound image of a human heart. The black square correspond to texture sample taken from a normal area (2) and an indicated infarcted area of myocardium (1).





(c)

Figure 7: a. Wavelet decomposition of normal myocardium, b. infarcted myocardium, and c. the arrangement of the four subbands (LL is low-frequency content of the original picture, LH gives the horizontal high frequencies, HL corresponds to vertical high frequencies and HH the high frequencies in both directions).

TABLE 1: Distance value (D) between normal and infarcted zones, Threshold value $\alpha = 2.7461$.

Patient	D
Infarcted Group	
P1	3.15040
P2	3.02680
P3	3.74600
P4	4.49660
P5	2.81280
P6	2.87220
P7	2.75430
P8	3.88320
P9	3.09640
P10	4.18970
Normal Group	
P11	0.49733
P12	0.98083
P13	1.06870
P14	1.71170
P15	0.39482

4. Conclusions

The following conclusions can be drawn from our studies:

- 1). Algorithm for texture analysis has advantage for detect difference image from echocardiography as normal myocardium and myocardium infarction.
- 2). Wavelet extension and co-occurrence matrix procedure approach is an effective method for application in similarity evaluation of texture images.

References

- [1]. S.G.Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation." IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, pp.674-693, July 1989.
- [2]. S.G.Mallat, "Multifrequency Channel Decomposition of Images and Wavelet Models." IEEE Transaction on Acoustics, Speech and Signal Processing, Vol.37, No.12, December 1989.
- [3]. I.Daubechies, "The Wavelet Transform, Time-Frequency Localization and Signal Analysis." IEEE Trans. on Information Theory, vol. 36, no.5, pp.961-1004, September 1990.
- [4]. A.Materka and M.Strzelecki, "Tecture Analysis Methods - A Review," Technical University of Lodz, Institute of Electronics, COST B11 report, Brussels 1998.
- [5]. M.Antonini, M.Barlaud, P.Mathieu, and I.Daubechies, "Image Coding Using Wavelet Transform." IEEE Trans. on Image Processing, vol. 1, no. 2, pp.205-220, April 1992.
- [6]. T.Chang, and C.C.Jay Kuo, "Texture Analysis and Classification with Tree-Structured Wavelet Transform." IEEE Trans. on Image Processing, vol. 2, no. 4, pp.429-441, October 1993.
- [7]. D.Dunn, and W.E. Higgins, "Optimal Gabor Filters for Texture Segmentation." IEEE Trans. on Image Processing, vol. 4, no. 7, pp.947-964, July 1995.
- [8]. Amara Graps, "An Introduction to Wavelet," IEEE Computational Sciene and Engineering, vol. 2, no. 2, Summer 1995.
- [9]. A.Mojssilović, M.V.Popović, A .N.Nešković, and A.D.Popović, "Wavelet Image Extension for Analysis and Classification of Infarcted Myocardial Tissue." IEEE Trans. on Biomedical Engineering, vol. 44, no. 9, pp.856-866, September 1997.
- [10]. A.L.Amer, A.Ertüzün, and A.Erçil, "An Efficient Method for Texture Defect Detection: Subband domain Co-Occurrence Matrices." Image and Vision Computing, Vol.18/6-7, pp.543-553, May 2000.
- [11]. Z.Shaohua, "Wavelet-Based Texture Retrieval and Modeling Visual Texture Perception." Thesis, Department of Electrical Engineering, National University of Singapore, 2000.
- [12]. T.Ojala, M.Pietikäinen, and T.Mäenpää, "Multiresolution Gray-Scale and Resolution Invariant Texture Classification with Local Binary Pattern," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp.971-987, July 2002.
- [13]. E.Chiu, J.Vaisey and M.S.Atkins, "Wavelet Based Space-Frequency Compression of Ultrasound Images," School of Engineering Science, School of Computing Science Simon Fraser University, Burnaby, BC, V5A 1S6, Canada, February, 2001.
- [14]. M.Partio, B.Cramariuc, M.Gabbouj and A.Visa, "Rock Texture Retrieval using Gray Level Co-occurrence Matrix." Tempere University of Tecnology, Tempere, Finland.

Study of Insulator to Withstand Switching Surges in Conversion Three to Six-Phase Transmission Line: Computer Simulation Analysis

Muhammad Irfan Jambak¹, Hussein Ahmad²

*Institute of High Voltage and High Current (IVAT)
Faculty of Electrical Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Bahru, Malaysia
Tel: +60-7-5535692, Fax: +60-7-5578150
E-mail: ijmuhammad2@siswa.utm.my¹
hussein@fke.utm.my²*

Abstract

Conversion of three to six-phase transmission line an alternative method to increase the present transmission line capability to meet the increasing electrical energy demand. In realizing this concept into actuality while maintaining same physical line dimensions and the same level of, the power capacity is more than 73% higher compare to three phase system. However there may exist in the conversion some impact will occur on the insulation of tower and substation equipment (considering no new items to be added) of the power system. This impact is associated with the stress due to switching phenomenon in the networks. This paper presents findings on the study of switching surges magnitudes on six-phase system by using EMTP/ATP. It was found that the magnitude of phase-to-phase switching surges of the converted system is much higher the parent system approximately 13% for phases spaced 120° and 65% for phases spaced 180°.

Keywords:

Transmission line, switching transient, over-voltages

1. Introduction

In recent years, rapid growth of Malaysia's economy has caused an increased on demand of electricity supply and load currents of transmission lines. Apart from this the national utility board has been corporatized to reduce government financial burden. To date, to cut cost on new line installations, instead of double circuit transmission lines as the main power transfer lines is used to fulfill the demand. However, the utility planner still need to anticipate the increasing demand of electrical power in advance since generation projects can have long lead times for future.

In the past [1], increase in power transmission capability has been accomplished by increasing system voltages. However, increasing of transmission operating voltage will produce strong electric field at ground level with possible biological aspect and environmental effects which necessitate large Right-of-Way (ROW). In consideration of the fundamental limits on power transfer capability in a restricted ROW led to the concept of increasing the number of phases in a transmission line system circuit as known as Multiphase system or High Phase Order (HPO) High Phase Order, is defined by number of voltages of equal magnitude, equally space in time [2]. For three phase system, this means three equal voltage spaced 120° in time. For a

6-phases this becomes 6 voltages spaced 30° and so on. HPO is a unique approach in increasing the power transmission capability of a overhead electric power transmission. Since Barthold and Barnes [3] was proposed HPO in 1972, the concept of six-phase transmission line (SPTL) has received great attention from researcher and it has been described in the literature in several paper and reports as alternative to increase the power transfer capability of existing three phase double circuit transmission line is the use of six-phase single circuit transmission line [4,5].

2. Impact of Phase Conversion

SPT is sometimes known as High Phase Order system is actually a system more than three-phase [7,8]. Phasor diagram of phase-to-phase and phase to ground is shown in Figure 1, and Figure 2 shows phase-ground-phase DGC triangle.

The equation of V_{line} and V_{phase} can derived by determining the resultant of DGC triangle in Figure 2 :

$$V_{CD} = 2 \times V_{CG} = 2 \times V_{CG} \cos \theta \quad (1)$$

Angle θ for adjacent phase-to-phase is 60° , it can simplified that

$$V_{line (adjacent)} = V_{CD} = 2 \times V_{phase} \cos 60^\circ \quad (2)$$

Hence,

$$V_{phase} = V_{line (adjacent)} \quad (3)$$

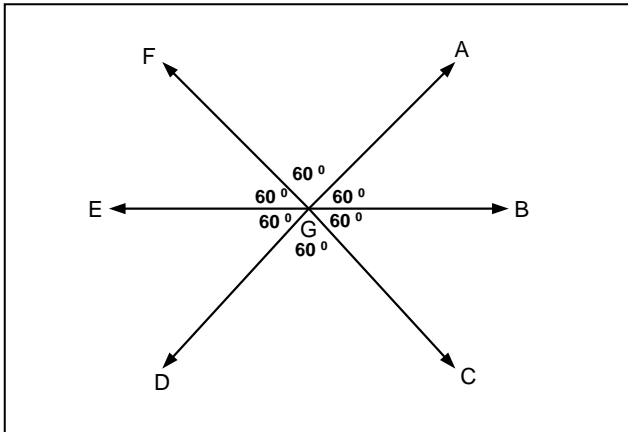


Figure 1. Phasor diagram of SPT system

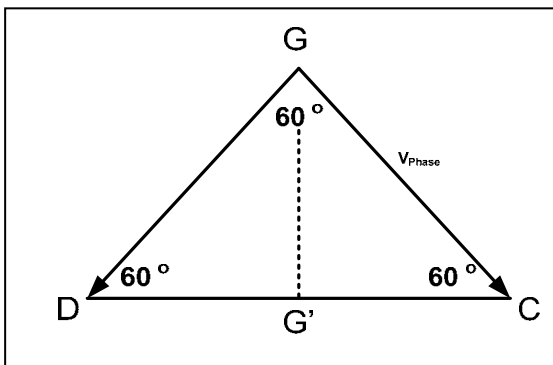


Figure 2. Phase-ground-phase triangle

Because $V_{phase (6 \text{ phase})}$ $\sqrt{3}$ higher than $V_{phase (3 \text{ phase})}$, hence, the main advantage of a six-phase transmission line it can carry up to 73% more electric power transfer capability compare to a three-phase system at the same operating voltage.

From Figure 1, the voltage system can be classified into four discrete voltages, i.e.: phase-to-ground voltage, between adjacent phase, between phases separated by one intermediate phase, and between opposite phases [10]. Within each group the voltages have identical magnitudes. In the group I the voltages are spaced 60° , in the group II and III the voltages are spaced 120° and 180° respectively. For example, when the six-phase transmission line is energize with a nominal phase-to-ground voltage of 93 kV, the phase-to-phase voltage will be 93 kV between adjacent

phases, 161 kV between phases 120° apart, and 186 kV between opposite phases or in the other hand voltage between adjacent phases is the same as the ground voltage, U , but the voltage between alternate phases is $\sqrt{3} U$ and between opposite phases is $2U$. So voltage stress on the insulators of six-phase mode will be substantially different from those in the three phase mode.

3. Switching Over-voltages

Over-voltages in power systems can be due to external and internal phenomena, such as faults, switching operation and lighting strokes to the tower or phase line. It is not practical to design the system equipment to withstand all type of over-voltages. In order to study insulation coordination it is necessary to analyze the different kinds of insulation levels for existing equipment before conversion three phase systems to six-phase system.

Switching over-voltages are resulted from operation of switching devices, either during normal conditions or as result of fault clearing. These transients have duration from tens to several hundreds of microsecond. The main operations that can be produce switching over-voltages are line energization and re-energization [6].

4. Modeling Requirements

Three phase double circuit line of Tenaga Nasional Berhad system has been chosen for the study. The chosen line is the 132kV three-phase double circuit transmission line between Gua Musang (GMSG) and Kuala Krai (KKRI), Kota Bahru Region, Kelantan which has a distance of 113.100 km in length. The one-line diagram of the system is shown in Figure 3.

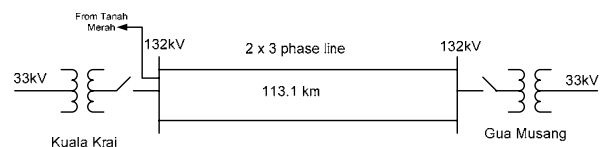


Figure 3. 132 kV double circuit line between KKRI and GMSG

The transmission line data required includes: transmission line conductor diameter, resistance per unit length, total length of transmission line, spacing between conductor on tower, shield wire, height of each conductor, sag to mid-span and tower dimensions. The selected transmission line model is distributed parameter models based on the traveling time and characteristic impedance of the line. The conductor used in the transmission line is called ACSR 300mm² for phase line and ACSR 80mm² for is used for earth wire. Two earth wires are used per tower, one on each side. Figure 4 shows phase and line arrangement at the

main intake of 132kV double circuit system between KKRI and GMSG.

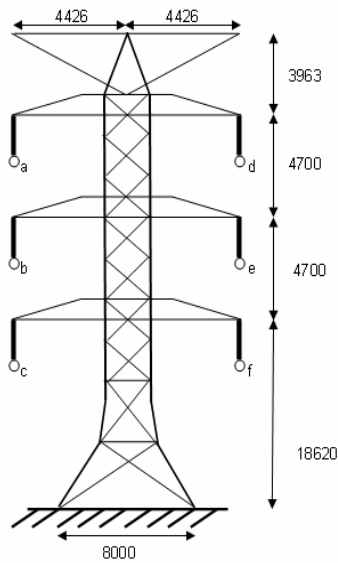


Figure 4. Phase conversion physical arrangement and tower dimension

5. Computer Simulation

The system configuration used to obtain the switching transient is illustrated in figure 5 where V_S is the supply voltage, CB is circuit breaker, and T is phase conversion transformer

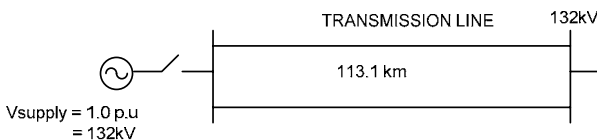


Figure 5. System Studied

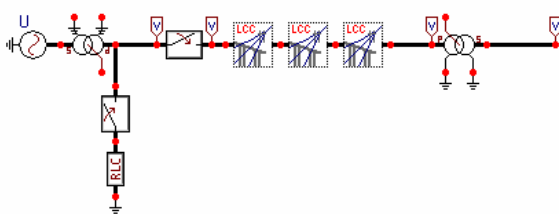


Figure 6. Three-Phase double circuit

Several modes of switching a six-phase line system is proposed i.e: single phase switching, two adjacent phases switching, three phases switching of line supplied by first transformer, and last six-phase switching. For phases “a” and phase “d” time open for circuit breaker (CB) is 0.03333 second and time closed is 10 second. For phases “b” and “e” time open for CB is 0.0361 and time closes is 10 second. For

phase “c” and “f” time open of CB is 0.0388 and time closed is 10 second. Figure 6 and Figure 7 shows studied model circuit by using ATPDraw for three-phase and six-phase respectively. In this model, 113 km transmission line is divided into three mid-spans. Voltage waveform at sending-end, after CB, each mid-span and receiving-end is observed.

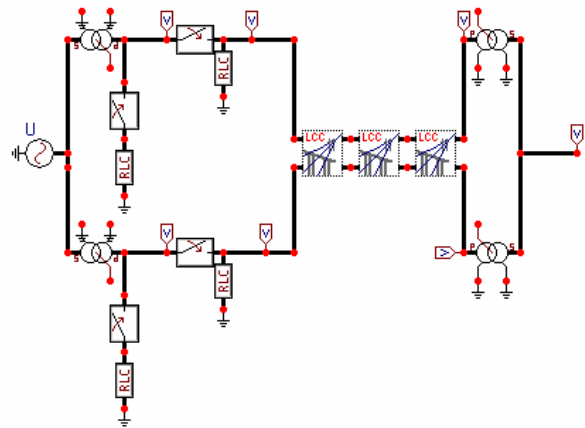


Figure 7. Six-Phase switching circuit

6. Simulation Result

Figure 8 through 10 shows switching transient simulation result of 132kV three phase double circuit system. Figure 8 typical voltage waveform obtained at upstream of circuit breaker, while figure 9 typical voltage obtained at bus bar.

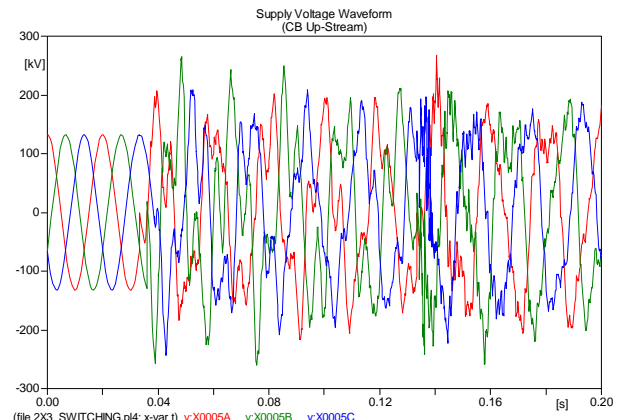


Figure 8. Three Phase Energizing voltage waveform

Table 1. Phase-to-Ground Switching Surge Magnitudes 132kV 2x3phase System

Probe Location	Phase A (p.u)		Phase B (p.u)		Phase C (p.u)	
	Cir 1	Cir 2	Cir 1	Cir 2	Cir 1	Cir 2
CB Up-Stream	2.0		1.98		1.79	
Sending-End	2.0		1.98		1.79	
Receiving-End	2.6		2.48		2.34	
37.6 km ^(*)	2.14	2.14	2.18	2.18	2.28	2.28
75.3 km ^(*)	2.36	2.36	2.27	2.27	2.30	2.30

Note:

(* distance from sending-end

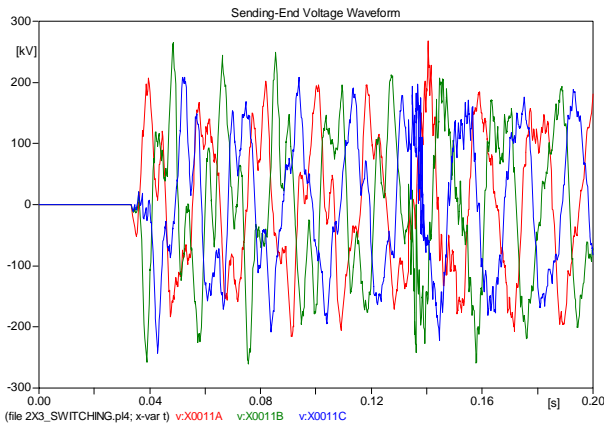


Figure 9. Sending-End voltage waveform after energizing

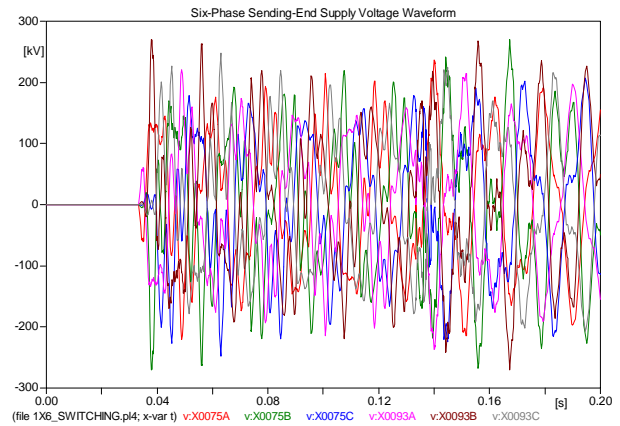


Figure 12. 6φ Sending-End voltage waveform after energizing

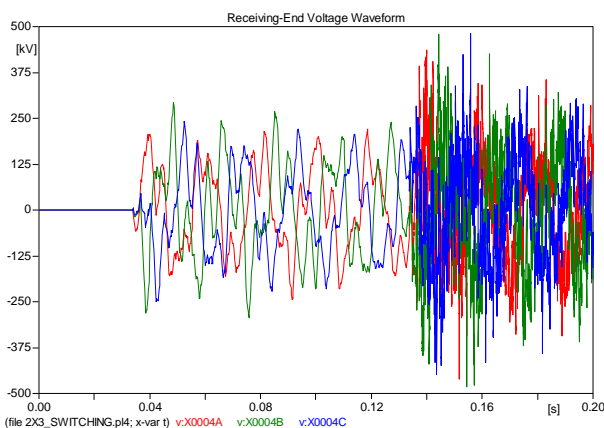


Figure 10. Receiving-End voltage waveform after energizing

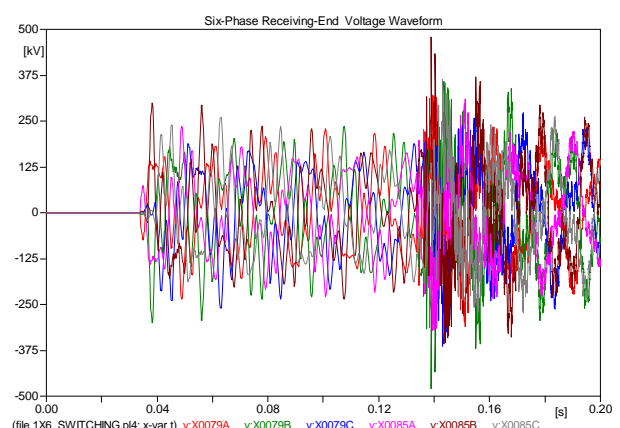


Figure 13. 6φ Receiving-End voltage waveform after energizing

Table 1 summarizes the highest value of maximum magnitude voltage obtained as described in figure 8 through 10 for the various locations.

Figure 11 through 15 shows switching transient simulation result of 132kV six-phase single circuit system with various switching condition and location. Figure 11 till figure 13 switching occurred at all line of six-phase line, while figure 14 and 15 are switching transient occurred only three phases (a-c-e) of six-phase line

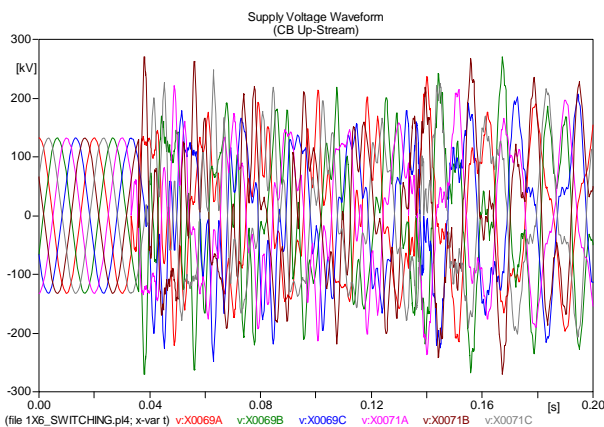


Figure 11. 6φ Energizing voltage waveform

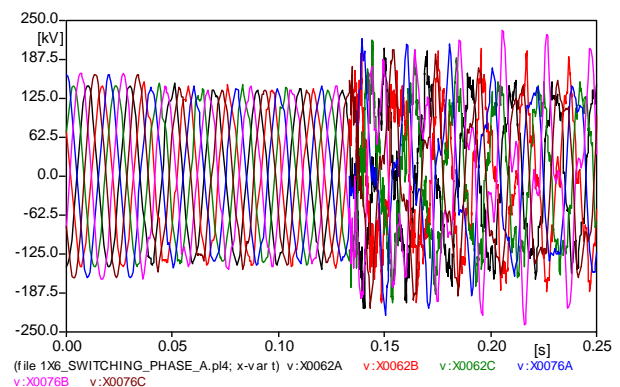


Figure 14. Sending-End voltage waveform after de-energizing phases a-c-e

In the series of simulations, line energization and re-energization are examined. Considering the above result, it can be concluded that some over-voltages are very severe. Circuit breaker and all equipment at substation need to be equipped with surge arrester.

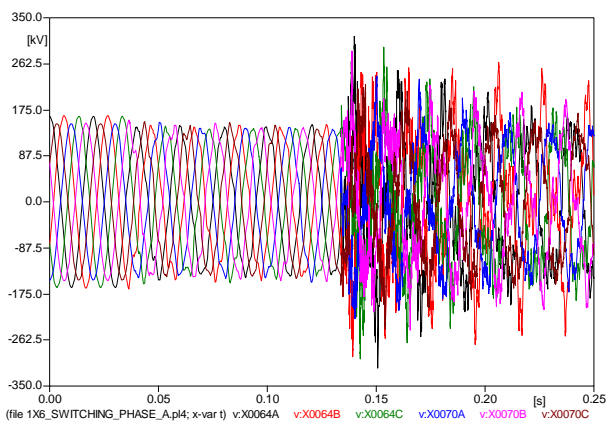


Figure 15. Receiving-End voltage waveform after de-energizing phases a-c-e

Table 2. Phase-to-Ground Switching Surge Magnitudes 132kV 1x6 phase System

Phase Switching Magnitude (p.u)						
Location	a	b	c	d	e	f
CB Up-Stream	1.79	1.79	2.04	2.04	1.87	1.87
Send-End	1.79	1.79	2.04	2.04	1.87	1.87
Rec-End	2.36	2.36	2.72	2.72	2.48	2.48
37.66 km (*)	1.98	1.98	2.27	2.27	2.47	2.47
75.32 km (*)	2.23	2.23	2.51	2.51	2.29	2.29

Note:

(* distance from sending-end)

Table 3. Phase-to-Ground Switching Surge Magnitudes 132kV 1x6 phase System (re-energization only at phase a-c-e)

Phase (p.u)	Location	
	Send-End	Rec-End
a	1.65	2.34
b	1.52	1.81
c	1.79	2.14
d	1.66	2.18
e	1.60	2.16
f	1.57	2.25

Table 4. Phase-to-Phase Switching Surge Magnitudes 132kV 1x6 phase System

Probe Location	3 Phase System	6 Phase System		
		Phase spaced 60°	Phase spaced 120°	Phase spaced 180°
CB Up-Stream	3.0	2.34	3.33	4.09
Sending-End	3.0	2.34	3.33	4.09
Rec- End	3.76	3.46	4.27	6.23
37.6 km(*)	3.06	2.96	3.38	5.0
75.3 km(*)	3.36	3.0	3.75	5.24

Note:

(* distance from sending-end)

3 phases re-energization (phase a-c-e) of six-phase line may cause impact switching over-voltages to other 3 phases of six phase line (phase b-d-f). Per unit value of these over-voltages at any location is lower compare to six-phase energization. To simplify the comparison switching surge magnitude between 3-phase system and 6-phase system, the maximums magnitude are in per unit of the respective normal operating voltages 132kV either phase-to-ground or phase-to-phase. The phase-to-ground switching surges for the 6-phase system little bit lower than 3-phase system. However, the phase-to-phase surge magnitudes of 6-phase system as described in Table 4 are significantly higher than 3-phase system, approximately 13% for phases spaced 120° and 65% for phases spaced 180°.

7. Conclusions

This paper reports the studies on the switching transients caused by energization and re-energization on six-phase power transmission line. The modeling and simulation is successfully implemented in ATP-EMTP software. Computer simulation studies are presented on a 132kV under 3 phase and six-phase system with various conditions and various location of point investigation. It is possible with these results to proceed and could be considered as a reference for experimental studies in the lab.

Acknowledgement

The authors wish to thank the Intensified of Research in Priority Areas (IRPA) authority for sponsoring this work. In addition, the authors are grateful to the Dean allowing us to publish the research findings, and to all the staff of Institut Voltan dan Arus Tinggi (IVAT), Faculty of Electrical Engineering, Universiti Teknologi Malaysia for their technical assistance.

References

- [1] M. Kanya Kumari, R. S. Shivakumara Aradhya, Channakeshava,(1992). "Multiphase Transmission System" ., CPRI Report, Published in CPRI News, July 1992, pp 5-7
- [2] James. R. S and Thomas L. H., (1998). "138kV 12-Phase as an Alternative to 345kV 3-Phase". IEEE Transaction on Power Delivery, Vol.13, No.4, October 1998
- [3] Venkata, S. S., et al., (1982). " EPPC – A Computer Program for Six Phase Transmission Line Design", IEEE Transaction on Power Apparatus and System, Vol. PAS-101, No. 7, July 1982, pp 1859-1869.
- [4] A.K. Mishra., A. Chandrasekaran., S.S. Venkata., (1995). "Estimation of Errors in the Fault Analysis of Six Phase Transmission Line Using Transposed

- Models". IEEE Transaction on Power Delivery, Vol.10, No.3, July 1995
- [5] N.B. Bhat., S.S. Venkata., W.C. Guyker., W.H. Booth., (1977). "Six-Phase (Multi-Phase) Power Transmission System: Fault Analysis". IEEE Transaction on Power Apparatus and Systems, Vol. PAS-96, No.3, 1977
- [6] CIGRE Working Group 13-02 Switching Surges Phenomena in EHV Systems, "Switching Overvoltages in EHV and UHV Systems with Special Reference to Closing and Re-closing Transmission Lines," Electra, Vol.30, pp70-122, 1973.
- [7] S.N. Tiwari., A.S. Bin Saroor., (1995). "An Investigation Into Loadability Characteristics of EHV High Phase Order Transmission Lines". IEEE Transaction on Power Systems, Vol.10, No.3, August 1995.
- [8] Kanya Kumari. M., Shivakumara Aradhya. R.S., Channakeshava., (1993). "Multiphase Transmission Lines and Their Feasibility". Presented on Workshop Trends in Electric Power Transmission Technology. 10-11 June 1993. CPRI

Design and Modeling of On-Chip Planar Capacitor for Radio Frequency Application

Mariyatul Qibthiyah Mohd Noor and Mazlina Esa*

*Microwave/RF and Antenna Research Group, Department of Radio Communication Engineering,
Faculty of Electrical Engineering, Universiti Teknologi Malaysia,
81310 UTM Skudai, Johor Darul Takzim, Malaysia.
Tel: +60-7-5535262, Fax: +60-7-5566272, Email: mazlina@fke.utm.my*

Abstract

The on-chip RF capacitor is one of the key components for RF integrated circuit designs such as filters and oscillators. This paper presents the investigations performed on the design of an interdigital capacitor configuration for RFICs. Geometry design variables include number of finger, finger length, finger width, finger gap, end gap, terminal width, strip thickness, substrate height, metal types and dielectric constant. Quality factor is the essential parameter as it is an index for the efficiency of a capacitor's performance. The physical model of an interdigital capacitor is first determined. Then, its equivalent lumped circuit is simulated. The effects of parameter variations on quality factor and capacitance value are discussed and shown. An optimised interdigital capacitor can be obtained through their performances.

Keywords:

Interdigital Capacitor, Quality Factor, Capacitance, Radio Frequency

1. Introduction

Many research activities are focusing on the possibilities to migrate the radio frequency (RF) circuit to CMOS technology. The on-chip planar capacitor is one of the major areas of such investigations. Advances in CMOS fabrications have resulted in deep submicron transistors with higher transit frequencies and lower noise figure. The advanced performance of MOSFET is attractive for high frequency circuit design in view of a system on-chip realization, where digital, mixed signal baseband and high frequency transceiver blocks would be integrated on a single chip. The ability to integrate RF circuit with other analog and logic circuits provides great demand in the future applications. Capacitors have become ubiquitous in analog integrated circuits particularly owing to the switched capacitor technique for realization of analog to digital and digital to analog data converters and discrete time filters. [1]-[4] This paper investigates the design of interdigital capacitor. Geometry design variables include number of finger, finger length, finger width, finger gap, finger end gap, terminal width, substrate height, metal types and dielectric constant.

Corresponding Author. E-mail: mazlina@fke.utm.my
Tel: +60-7-5535262, Fax: +60-7-5566272

The optimum design of interdigital capacitor can be identified through the development of Q -factor contour

plots. The research involves mathematical computation using MathCAD [5] and electromagnetic simulation using SonnetLite Plus [6]. Previous work are available in [7]-[18].

2. Computations Using MathCAD

The suitable geometry parameter value can be obtained from the mathematical analysis.

2.1 Solving of Formulation

From the desired design specification, the relevant formulations can be solved.

Firstly, from the formulation of quality factor, Q , set the desired Q of 240. Then, quality factor due to the conductor losses, Q_c , can be obtain as

$$Q_c = \frac{Q Q_d}{Q_d} = 6000 \quad (1)$$

Hence, the quality factor due to dielectric losses, Q_d , can be obtain from

$$Q_d = \frac{1}{\tan \delta} = 250 \quad (2)$$

where $\tan \delta$ is the dielectric loss tangent. Then, the equation of Q_c is given by [7]

$$Q_c = \frac{l}{\omega CR} \tag{3}$$

From equation (3), with $C = 0.2$ pF and $f = 5$ GHz, the value of R is computed to be 0.0265. The equation of series resistance, R is given by the formulation [7]

$$R = \frac{4lR_s}{3w} \tag{4}$$

where R_s is surface resistivity and is given by [7]

$$R_s = \sqrt{\rho\pi\mu_0 f} \tag{5}$$

A value of resistivity of the interconnect material, ρ , is set as $1.7 \times 10^{-8} \Omega\text{m}$ [8]. The permeability, μ_0 , is $4\pi \times 10^{-7} \text{H/m}$. Therefore surface resistivity, R_s becomes $2.5906 \times 10^{-7} \Omega$. From equation (4), the relationship between finger length, l , and capacitor width, w , can be expressed as

$$\frac{l}{w} = 1.08 \tag{6}$$

with w given by:

$$w = xn + s(n - 1) \tag{7}$$

Substituting equation (7) into (6) gives :

$$l = 1.08[n(x + s) - s] \tag{8}$$

The next step is to determine the dimensions of the interdigital capacitor.

2.2 Graphical Approximation Method

The graphical approximation method has been chosen in setting the suitable values for the design variables parameter. For the first analysis, l and s are fixed to 0.28 mm and 0.02 mm, respectively. From equation (8), suitable combination values for finger width, x , and number of finger, n , have been obtain as shown in Figure 1 and Table 1.

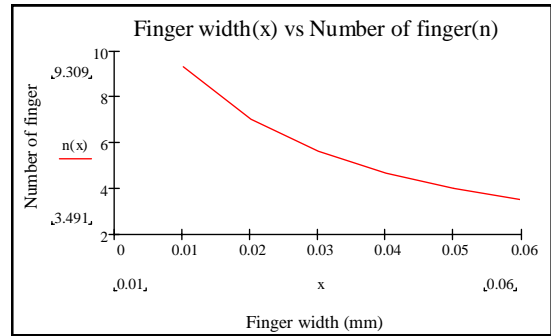


Figure 1: x versus n when l and s are fixed

Table 1: Suitable combination values of x and n

Geometry Variable	Parameter Value			
x , mm	0.01	0.02	0.03	0.04
N	9.31	6.98	5.68	4.65

For the second analysis, the finger width, x , is fixed to 0.02 mm while the value of l remains. Therefore, suitable combination values for finger spacing, s , and number of finger, n , have been produced as shown in Figure 2 and Table 2.

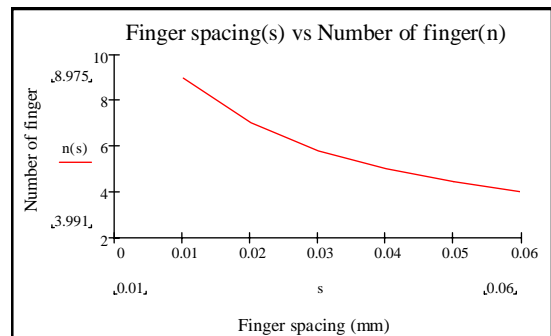


Figure 2: s versus n when l and x are fixed

Table 2: Suitable combination values for s and n

Geometry Variable	Parameter Value			
s , mm	0.01	0.02	0.03	0.04
n	8.98	6.98	5.78	4.98

This third analysis is to determine the suitable combination values for n and l when both x and s are fixed to 0.02 mm. The behaviour is shown in Figure 3 and Table 3.

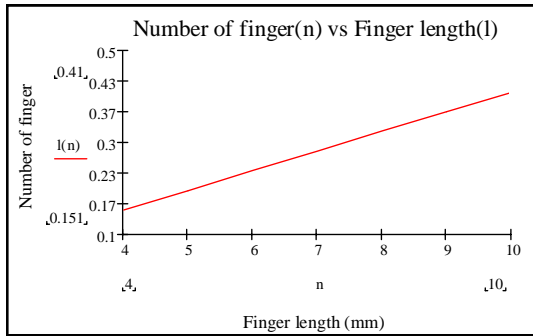


Figure 3: n versus l when s and x are fixed

Table 3: Suitable combination values for n and l

Geometry Variable	Parameter Value		
n	6	7	8
l , mm	0.23	0.28	0.32

From the mathematical analysis, five interdigital capacitor designs have been created. All related variable parameters for these designs are shown in Table 4. Design 1 is set as the reference design.

Table 4: Design geometries of the interdigital capacitor

	x , mm	s , mm	l , mm	n , mm
Design 1	0.02	0.02	0.28	7
Design 2	0.04	0.02	0.28	5
Design 3	0.02	0.04	0.28	5
Design 4	0.02	0.02	0.24	6
Design 5	0.02	0.01	0.28	9

3. Simulation Result

3.1 SonnetLite Plus

By using an electromagnetic simulation software, the characterization of interdigital capacitor permits more flexibility during the design process. The effect of interdigital capacitor's geometry variations can be more easily analyzed.

3.2 Effect on Quality Factor and Capacitance

3.2.1 Design Comparison

For this analysis, Design 4 which has the smallest dimension produces the highest Q_{max} of 249.7868 and lowest capacitance value of 0.06065 pF. However, Design 1 is chosen as the basic design because the design fulfills the desired specification such as Q factor of more than 240, capacitance between 0.08 and 0.2 pF, and operating frequency from 2 to 5 GHz.

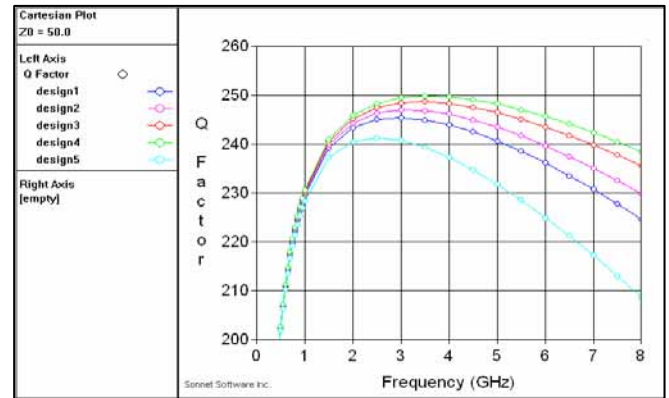


Figure 4: Effect on Q_{max} for different designs of interdigital capacitor

3.2.2. Number of Fingers

For the second analysis, several designs were created with different number of fingers i.e. 4, 5, 6, 8, 9 and 10, compared to Design 1 which has 7 fingers. The capacitance value increases with increasing number of fingers as shown in Figure 5. However, the Q factor is inversely proportional to n . As shown in Figure 6, smallest number of finger produces highest Q factor value. Approximately 1.73 % decrement of Q_{max} occurs when n increases from 7 to 10.

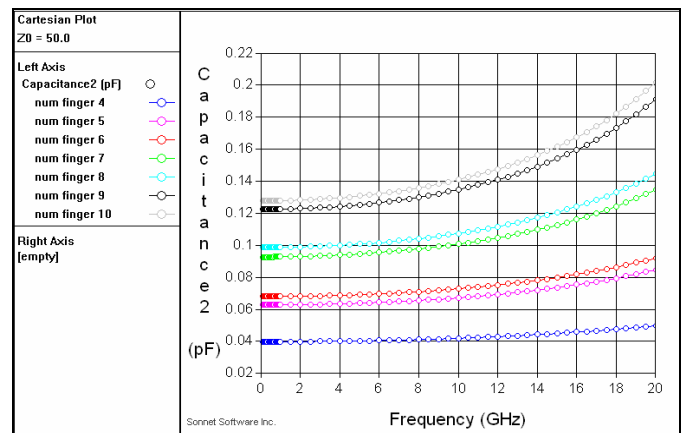


Figure 5: Effect on capacitance, C , value for different n .

3.2.3 Types of Metal

Several designs have been created using different metals to compare with Design 1 which uses copper as the metal layer. When better conductors are used, or metals with higher conductivity, the metallization loss is reduced but increased the Q factor value as shown in Figure 7 and Table 5. Different types of metals give low impact to the capacitance curve.

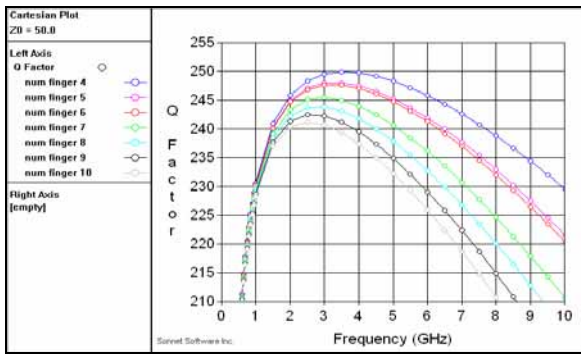


Figure 6: Effect on capacitance, C value for different n .

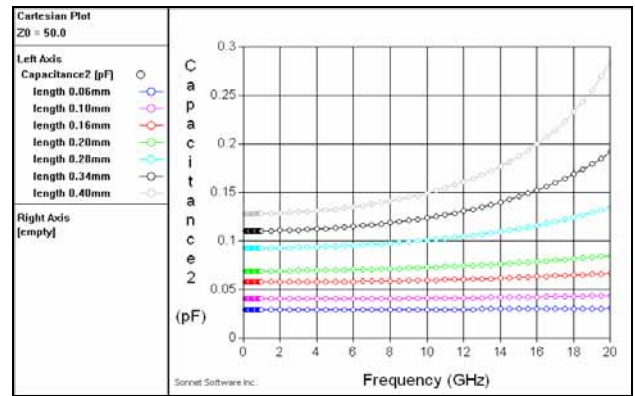


Figure 9: Effect on capacitance, C value for different l .

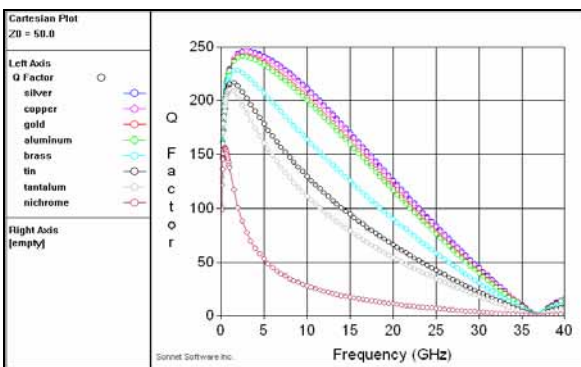


Figure 7: Effect on Q_{max} value for different types of metal

Table 5: Q_{max} value of different types of metal [8]

Metal Type	Conductivity, S/m	Q_{max}	Frequency at Q_{max} , GHz
Silver	6.17×10^7	245.94	3.0
Copper	5.8×10^7	245.40	3.0
Gold	4.09×10^7	241.90	2.5
Aluminum	3.72×10^7	240.85	2.5
Brass	1.57×10^7	228.03	2.0
Tin	8.7×10^6	216.60	1.5
Tantalum	6.45×10^6	209.82	1.0
Nichrome	1×10^6	156.52	0.45

3.2.4 Finger length

Several designs using $l = 0.06$ mm, 0.10 mm, 0.20 mm, 0.34 mm and 0.40 mm were created to compare with Design 1 which uses $l = 0.28$ mm. Capacitance value increases when finger length increases. The design with $l = 0.06$ mm

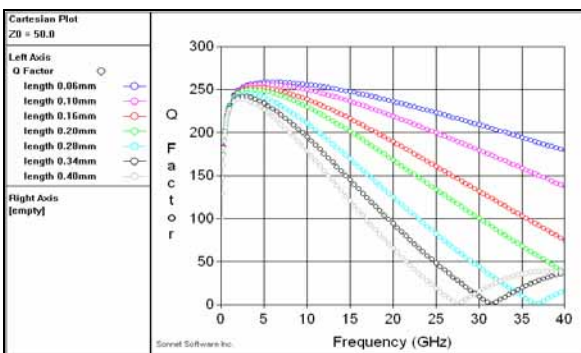


Figure 8: Effect on Q_{max} value for different l . produces highest Q_{max} of 258.4685 which is approximately 5.32 % increment compared to Design 1. The results are shown in Figures 8 and 9. Shorter finger length reduces the series resistance that eventually leads to higher Q factor.

4. Further Discussion and Conclusion

From the analysis, it can be summarized that Q_{max} increases with decreasing number of fingers, smaller finger length and higher metal's conductivity. In order to achieve high Q_{max} , finger length plays an important role because it affects more compared to other parameters. By using a small finger length of 0.06 mm, approximately 5.32 % increment of Q_{max} is observed. However, designs of small finger length produce low capacitances. As a result, a 0.28 mm finger length is an ideal value as the Q_{max} can reach 245.4057 at 3 GHz and at the same time produces capacitance, C of 0.093380 pF, which is considered as medium capacitance value.

The capacitance, C increases with increasing number of fingers and larger finger length. The metal's conductivity gives low impact to the capacitance. From the analysis, Design 5 with nine number of fingers, 0.28 mm finger length, 0.02 mm finger width and 0.01 mm finger spacing, produces the highest capacitance, C of 0.159 pF. The capacitance, C , value increases by approximately 70 % compared to the original design. However, larger finger length and more number of fingers will consume larger cost to the RFIC design.

From the investigations performed, it can be concluded that interdigital capacitor produces high quality factor in the range of 240 and medium capacitance value. The material

properties and geometrical parameters for an interdigital capacitor are very important in order to optimize the quality factor and capacitance. The full investigation is reported in [19].

References

- [1] T. Ytterdal, Y. Cheng and T. A. Fjeldly, *Device Modeling for Analog and RF CMOS Circuit Design*. Wiley 2003
- [2] F. P. Casares-Miranda, P. Otero, E. Marquez-Segura and C. Camacho-Penalosa "Wire Bonded Interdigital Capacitor," IEEE Trans MICROWAVE AND WIRELESS COMPONENTS, VOL 15, No.10, 2005
- [3] S. Hontsu, S. Mine, H. Nishikawa, M. Nakamori, A. Fujimaki, M. Inoue, A. Maehara, T. Kawai, "Study of Mechanically Tunable Superconducting Microwave Filter Using Lumped Elements," IEEE Trans. on Applied Superconductivity, Vol 3, No. 2, 2003
- [4] Leo G. Maloratsky, *Passive RF & Microwave Integrated Circuits*. Elsevier 2004
- [5] <http://www.mathsoft.com>
- [6] <http://www.sonnet.usa>
- [7] W. F. Mullin, *ABC's of Capacitors*, Howard W. Sams & Co, 1966
- [8] R. Brown "RF/Microwave Hybrids: Basics, Materials and Processes," Kluwer Academic Publishers 2002
- [9] Ian Sinclair, *Passive Components for Circuit Design*, Newnes, 2001
- [10] Gary. D. Alley, "Interdigital Capacitor and Their Application to Lumped-Element Microwave Integrated Circuits", IEEE Trans. Microwave Theory Tech, Vol MTT-18 1970
- [11] R. Esfandiari, D. W. Maki, M. Siracusa, "Design of Interdigital Capacitor and Their Application to GaAs Filters", IEEE Trans. Microwave Theory Tech, Vol MTT-31, 1983
- [12] E. Pattenhaul, H. Kapusta, A. Wisgerber, H. Mempe, J. Luginsland and I. Wolff, "CAD Models of Lumped Elements on GaAs up to 18GHz", IEEE Trans. Microwave Theory Tech, Vol MTT-36, 1988
- [13] F. Aryanfar and K. Sarabandi "Characterization of Semilumped CPW Elements for Millimeter-Wave Filter Design," IEEE Trans. Microwave Theory Tech Vol 53, No. 4, 2005
- [14] Inder Bahl, Prakash Bhartia, *Microwave Solid State Circuit Design*. Wiley 2003
- [15] S. S. Gevorgian, T. Martinsson, Peter L, J. Linner and E. L. Kollberg, "CAD Models for Multilayered Substrate Interdigital Capacitors," IEEE Trans. Microwave Theory Tech, Vol 44, No. 6, 1998
- [16] K. Chang, Inder Bahl and Vijay Nair, *RF and Microwave Circuit and Component Design for Wireless Systems*. Wiley 2002
- [17] Queennie S. I. Lim, Albert V. Kordesch and Richard A. Keating, "Performance Comparison of MIM Capacitors and Metal Finger Capacitors for Analog and RF Application," Proc. of 2004 RF and Microwave Conference, Shah Alam, Selangor, Malaysia
- [18] I. Kneppo and J. Fabian, *Microwave Integrated Circuits*, Chapman & Hall, 1994
- [19] Mariatul Qibthiah Mohd Noor, "Design and Modeling of On-Chip Planar Capacitor for Radio Frequency Application", Masters thesis, unpublished dissertation, Universiti Teknologi Malaysia, 2006.

Elimination of Mismatched Imbalance Vibration of Active Magnetic Bearing System Using Full Order Sliding Mode Controller

Husain A.R.¹, Ahmad M.N.¹ and Mohd Yatim A.H.¹

¹Faculty of Electrical Engineering,
Universiti Teknologi Malaysia,
81310, Skudai, Johor Bahru, Malaysia
Tel: +60-7-5535894, Fax: +60-7-5544272, E-mail: rashid@fke.utm.my

Abstract

This paper deals with the method to eliminate the vibration inherited in Active Magnetic Bearing (AMB) System due to mass imbalance of the rotor. The imbalance generates variable frequencies of sinusoidal disturbance forces that causes vibration in the system. The mathematical model of the AMB system derived is represented in state variable form and it is also shown that the system suffers the mismatched condition due to this disturbance. The method proposed to overcome the vibration effect is based on full order sliding mode control. In the design process, firstly, a sliding surface is designed that ensures the system is robust to uncertainties once it slides on the surface. Then, a controller that guarantees the system reaches and maintain on the sliding surface is proposed in which Lyapunov stability condition is satisfied. Several simulation result are obtained to evaluate the proposed method. The results show that the proposed controller is able to eliminate the vibration and the robust stability is also achieved.

Keywords:

Sliding Mode Control, Active Magnetic Bearing, Mismatched Rotor Imbalances, Vibration.

1. Introduction

An active magnetic bearing (AMB) system is a collection of electromagnets used to suspend and object and stabilization of the system is performed by feedback control. The system is composed of a floating mechanical rotor and electromagnetic coils that provide the controlled dynamic force. Due to this non-contact operation, AMB system has many promising advantages for high-speed and clean-environment applications. Moreover, adjustable stiffness and damping characteristics also make the system suitable for elimination of system vibration. Although the system is complex and considered an advance topic in term of its structural and control design, the advantages it offers outweigh the design complexity. A few of the AMB applications that receive huge attentions from many research groups around the globe are the flywheel energy, turbo molecular pump, Left Ventricular Assist Device (LVAD) and artificial heart [1].

Other than the gyroscopic effect, the vibration due to mass imbalance of the rotor mass has caused instability of the AMB system and this has imposed great challenge to eliminate the effect. This is due to the fact that the magnitude and frequency of sinusoidal disturbance depends on the rotating speed of the rotor. Also, from the model of AMB system, it can be seen that the vibration term is mismatched with the control input. Many methods have been proposed to overcome the effect such as Q-

parameterization [4] [5] and notch filter design [6]. However, the methods are designed at a few rotational speed. In this work, full order sliding mode control (FOSMC) technique will be used to eliminate the vibration effect by performing imbalance compensation where the input voltage will be generated in such a way that the rotor is rotating along its axis of geometry. The method is used due to its robustness towards system uncertainties and parameter variation.

This paper will be organized as follows: In section 2, the derivation of the mathematical modeling of cylindrical AMB system will be covered. In section 3, the synthesis of the FOSMC to remove the mismatched imbalance vibration will be shown. Then, the simulation and the result is discussed in section 4. Finally, the work shown this paper is summarized in the conclusion and a few of future works are outlined.

2. Mathematical Model of Active Magnetic Bearing System

The cross section view of the cylindrical AMB used for the derivation of the mathematical model is shown in Figure 1 below. Four of the voltage supplies, e_{11} , e_{12} , e_{13} and e_{14} are the top and bottom coils for both left and right that control the vertical motion of the rotor while the voltage supplies for the horizontal motions are e_{13} , e_{14} , e_{r3} and e_{r4} . (not shown in Figure 1).

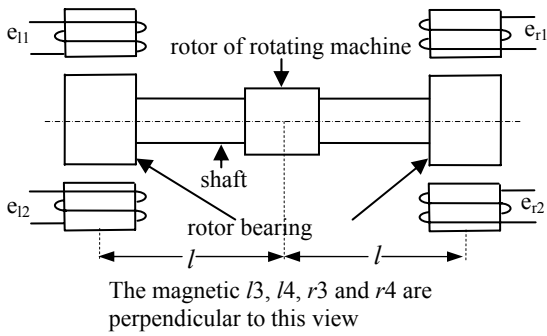


Fig. 1: Cross section view of cylindrical magnetic bearing from z-plane

2.1 Equation of motions

Assuming the rotor is rigid floating body, the fundamental equation of motions for the rotor for four degree of freedom are [2][3]

$$\begin{aligned} \ddot{y}_o &= \frac{1}{m} [\alpha y_o + (f_{l3} \quad f_{l4} + f_{r3} \quad f_{r4}) + f_{dy}] \\ \ddot{z}_o &= \frac{1}{m} [\alpha z_o + (f_{l2} \quad f_{l1} + f_{r2} \quad f_{r1}) + f_{dz} + mg] \\ \ddot{\theta} &= \frac{pJ_x}{J_y} \dot{\psi} + \frac{l}{J_y} ((f_{l1} \quad f_{l2} + f_{r2} \quad f_{r1}) + f_{d\theta}) \\ \ddot{\psi} &= \frac{pJ_x}{J_y} \dot{\theta} + \frac{l}{J_y} ((f_{l3} \quad f_{l4} + f_{r4} \quad f_{r3}) + f_{d\psi}) \end{aligned}$$

where

- m mass of the rotor
- J_x moment of inertia around x-axis,
- J_y moment of inertia around y-axis
- α radial eccentricity coefficient,
- β axial eccentricity coefficient.

Notice that the movement in x-axis is not included since for cylindrical magnetic bearing, axial motion is uncoupled to other motions and controlled separately. Also, $f_{l1}, f_{l2}, f_{l3}, f_{l4}, f_{r1}, f_{r2}, f_{r3}, f_{r4}$ are the nonlinear magnetic forces produced by the stator coils and exerted on the rotor while $f_{dy}, f_{dz}, f_{d\psi}, f_{d\theta}$ are the sinusoidal disturbance forces caused by the mass imbalance and given by the following equations [5]:

$$\begin{aligned} f_{dy} &= m_o \epsilon p^2 \cos(pt + \kappa) \\ f_{dz} &= m_o \epsilon p^2 \sin(pt + \kappa) \\ f_{d\theta} &= \frac{(J_y \quad J_x)}{l} \tau p^2 \cos(pt + \lambda) \\ f_{d\psi} &= \frac{(J_y \quad J_x)}{l} \tau p^2 \sin(pt + \lambda) \end{aligned} \tag{2}$$

where

- m_o mass of unbalance
- r radial distance the unbalance mass from rotor center
- ϵ, τ static and dynamic imbalance
- κ, λ initial values

2.2 Electromagnetic Equations

There are two ways to represent the electromagnetic dynamics which are current based or flux based. Each representation has its advantages [7] and for this work, current-to-force relation is used as shown by equation (3) below in which i_j is the coil current and g_j is the total gap length.

$$f_j = k \left(\frac{i_j}{g_j} \right)^2, \quad j = l1, l2, l3, l4, r1, r2, r3, r4. \tag{3}$$

and k is a constant.

The coil voltage, e_j , which is the input to the system can be expressed in term of i_j , coil inductance L and coil resistance, R . The equation that represents this relation is as follows:

$$e_j = L \frac{di_j}{dt} + Ri_j, \quad j = l1, l2, l3, l4, r1, r2, r3, r4. \tag{4}$$

2.3 Linearization

After integrating equations (1), (2), (3) and (4) above to form a complete cylindrical AMB system, 16 nonlinear equation are formed with 16 states. In order to synthesize the controller that can eliminate the vibration effect, the system is treated to operate around its nominal value with some deviations from this nominal operating conditions. Let F_{oj}, I_{oj}, G_{oj} and E_{oj} be the nominal values of the force, coil current, gap length and electromagnet voltage supplies of the j th electromagnet respectively. The deviation value of these quantities are f_j', i_j', g_j' and e_j' . For the deviation value of the force, the partial derivatives of equation (3) is calculated as follows:

$$\frac{\partial f}{\partial g} = k i^2 \left(-\frac{2g}{g^4} \right) = -\frac{2ki^2}{g^3} \tag{5i}$$

$$\frac{\partial f}{\partial i} = \frac{2ki}{g^2} \tag{5ii}$$

The quantities derived above is the linearization constants and assumed to be the same for all the electromagnet coils. Thus the quantities can be named as follows:

$$d_j = -\frac{2kI_o^2}{G_o^3}, \quad c_j = \frac{2kI_o}{G_o^2} \tag{5iii}$$

Then, the following equations can be written for the systems,

$$f_j = F_{oj} + f_j', \quad i_j = I_{oj} + i_j', \quad g_j = G_{oj} + g_j', \quad e_j = E_{oj} + e_j' \tag{6}$$

where

$$f_j' = c_j i_j' + d_j g_j', \quad e_j' = L \frac{di_j'}{dt} + R i_j'. \quad (7)$$

For cylindrical horizontal magnetic bearing,

$$\begin{aligned} d_{11} = d_{r1}, d_{12} = d_{r2}, d_{13} = d_{14} = d_{r3} = d_{r3}, \\ c_{11} = c_{r1}, c_{12} = c_{r2}, c_{13} = c_{14} = c_{r3} = c_{r3}. \end{aligned} \quad (8)$$

Also, the system is controlled in such a way that,

$$e_{11}' = -e_{12}', e_{r1}' = -e_{r2}', e_{13}' = -e_{r3}', e_{14}' = -e_{r4}', \quad (9)$$

so that the the currents in the coils and force terms will be as follows:

$$i_{11}' = -i_{12}', i_{r1}' = -i_{r2}', i_{13}' = -i_{r3}', i_{14}' = -i_{r4}', \quad (10)$$

$$\begin{aligned} f_{11}' - f_{12}' &= (c_{11} + c_{12})i_{11}' + (d_{11} + d_{12})g_{11}' \\ f_{r1}' - f_{r2}' &= (c_{r1} + c_{r2})i_{r1}' + (d_{r1} + d_{r2})g_{r1}' \\ f_{13}' - f_{14}' &= 2c_{13}i_{13}' + 2d_{13}g_{13}' \\ f_{r3}' - f_{r4}' &= 2c_{r3}i_{r3}' + 2d_{r3}g_{r3}' \end{aligned} \quad (11)$$

By using gap deviation expression found in [2] and treating $i_y = i_{13}' + i_{r3}'$, $i_z = i_{11}' + i_{r1}'$, $i_\theta = i_{11}' - i_{r2}'$, $i_\psi = i_{13}' - i_{r3}'$, the new states equation of the AMB system is shown below.

$$\begin{aligned} \ddot{y}_o &= \frac{1}{m} [(\alpha \quad 4d_{13})y_s + 2c_{13}i_y + f_{dy}] \\ \ddot{z}_o &= \frac{1}{m} [(\alpha \quad 2d_{112})z_s \quad c_{112}i_z + f_{dz}] \\ \ddot{\theta} &= \frac{pJ_x}{J_y} \dot{\psi} + \frac{1}{J_y} (d_{112} (2l\theta) + c_{112}i_\theta + f_{d\theta}) \\ \ddot{\psi} &= \frac{pJ_x}{J_y} \dot{\theta} + \frac{1}{J_y} (2d_{13} (2l\psi) + 2c_{13}i_\psi + f_{d\psi}) \end{aligned} \quad (12)$$

Also, from (9) and (10) the following new controlled input voltages can be established:

$$\begin{aligned} e_y &= e_{13}' + e_{r3}' = L \frac{di_y}{dt} + R i_y, \\ e_z &= e_{11}' + e_{r1}' = L \frac{di_z}{dt} + R i_z, \\ e_\theta &= e_{11}' - e_{r1}' = L \frac{di_\theta}{dt} + R i_\theta, \\ e_\psi &= e_{13}' - e_{r3}' = L \frac{di_\psi}{dt} + R i_\psi. \end{aligned} \quad (13)$$

2.4 AMB model in state-space representation

From the equations derived in previous subsection, the system equation now can be converted into state variable form as follows,

$$\begin{aligned} \dot{\mathbf{X}}_1 &= \mathbf{A}_y \mathbf{X}_1 + 0 \mathbf{X}_2 + 0 \mathbf{X}_3 + 0 \mathbf{X}_4 + \mathbf{U}(t) + \mathbf{d}(t) \\ \dot{\mathbf{X}}_2 &= 0 \mathbf{X}_1 + \mathbf{A}_z \mathbf{X}_2 + 0 \mathbf{X}_3 + 0 \mathbf{X}_4 \\ \dot{\mathbf{X}}_3 &= 0 \mathbf{X}_1 + 0 \mathbf{X}_2 + \mathbf{A}_\theta \mathbf{X}_3 + \mathbf{A}_{\theta\psi} \mathbf{X}_4 \\ \dot{\mathbf{X}}_4 &= 0 \mathbf{X}_1 + 0 \mathbf{X}_2 + \mathbf{A}_{\psi\theta} \mathbf{X}_3 + \mathbf{A}_\psi \mathbf{X}_4 \end{aligned}$$

where

$$\begin{aligned} \mathbf{X}_1 &= [\dot{y} \quad \ddot{y} \quad i_y]^T, \quad \mathbf{X}_2 = [\dot{z} \quad \ddot{z} \quad i_z]^T, \\ \mathbf{X}_3 &= [\dot{\theta} \quad \ddot{\theta} \quad i_\theta]^T, \quad \mathbf{X}_4 = [\dot{\psi} \quad \ddot{\psi} \quad i_\psi]^T. \end{aligned}$$

The components for the matrices are given as follows:

$$\begin{aligned} \mathbf{A}_y &= \begin{bmatrix} 0 & 1 & 0 \\ \frac{\alpha}{m} & \frac{4d_{13}}{m} & 0 \\ 0 & 0 & \frac{2c_{13}}{R/L} \end{bmatrix}, \\ \mathbf{A}_z &= \begin{bmatrix} 0 & 1 & 0 \\ \frac{\alpha}{m} & \frac{2d_{112}}{m} & 0 \\ 0 & 0 & \frac{c_{112}}{R/L} \end{bmatrix}, \\ \mathbf{A}_\theta &= \begin{bmatrix} 0 & 1 & 0 \\ \frac{2d_{112}}{m_1} & 0 & \frac{c_{112}}{lm_1} \\ 0 & 0 & R/L \end{bmatrix}, \\ \mathbf{A}_\psi &= \begin{bmatrix} 0 & 1 & 0 \\ \frac{4d_{13}}{m_1} & 0 & \frac{2c_{13}}{lm_1} \\ 0 & 0 & R/L \end{bmatrix}, \\ \mathbf{A}_{\theta\psi} &= \text{diag}(0, -\frac{pJ_x}{J_y}, 0), \quad \mathbf{A}_{\psi\theta} = \text{diag}(0, \frac{pJ_x}{J_y}, 0), \\ \mathbf{U}(t) &= [e_y \quad e_z \quad e_\theta \quad e_\psi]^T, \quad b_y = b_z = b_\theta = b_\psi = \text{diag}(0,0,1/L), \\ d_y &= d_z = \text{diag}(0, 1/m, 0), \quad d_\theta = d_\psi = \text{diag}(0, 1/m_1, 0), \\ m_1 &= J_y / l^2 \quad \text{and} \quad \mathbf{d}(t) = [f_{dy} \quad f_{dz} \quad f_{d\theta} \quad f_{d\psi}]^T. \end{aligned}$$

From this state equation, it can be seen that the system can be decoupled into 3 subsystems with 2 SISO systems that govern the y and z motions and 1 MIMO system that describe both θ and ψ motions. The controller derived in the next section will be based on these 3 subsystem in which

the vibration components in each subsystem will be eliminated.

3. Full Order Sliding Mode Controller (FOSMC) Design

The state equation (14) above can be written as

$$\dot{X}(t) = AX(t) + BU(t) + f(t) \tag{15}$$

where $f(t)$ is the vibration term with the mismatched condition i.e. $rank [B, f(t)] \neq rank [B]$ in which this term is not in phase with the system input. Thus, the proposed controller should be robust enough to overcome the mismatched vibration term to ensure stability of the system performance can be achieved.

Let the PI Sliding Surface be defined as [8]:

$$s(t) = CX(t) - \int_0^t [CA - CBK] X(\tau) d\tau \tag{16}$$

where $C \in^{m \times n}$ and $K \in^{n \times n}$ are constant matrices.

The matrix C is chosen such that $CB \in^{m \times m}$ is non-singular while the matrix K satisfies $\lambda_{\max}(A + BK) < 0$. This condition will guarantee that all desired poles are located in the left-hand plane to ensure system stability. It is well known that if the system is able to enter the sliding mode, $s(t) = 0$. Then, the equivalent control, U_{eq} can be determined by letting $\dot{s}(t) = 0$ and substituting the equation into (15) producing

$$U_{eq} = KX(t) - (CB)^{-1}Cf(t) \tag{17}$$

Substituting (17) into system equation (15) gives the equivalent dynamic equation of the system when sliding on the surface as:

$$\dot{X}(t) = (A + BK)X(t) + \{I_n - B(CB)^{-1}C\}f(t) \tag{18}$$

Notice that the system will be asymptotically stable if vibration term is matched eliminating the second term in (18). Due to this mismatched condition, the following theorem is introduced.

Theorem 1: If $\|\tilde{F}(t)\|$ $\beta_1 = \|I_n - B(CB)^{-1}C\| \beta$, the uncertain system equation (18) is boundedly stable when sliding surface.

Proof: See [9].

For the controller design that can drive the system states to slide on sliding surface and ensure system remains boundedly stable thereafter, the controller of the following form is proposed:

$$U = U_{eq} + U_{nl} \tag{19}$$

where the total control effort U is composed of equivalent control, U_{eq} derived in (17) and nonlinear control, U_{nl} . The equivalent control is stabilizing the system on sliding surface and the nonlinear control is the nonlinear feedback control that suppress the disturbance and forcing the system to switch on the designed surface. Thus, the following control law is proposed:

$$U(t) = KX(t) - (CB)^{-1}Cf(t) - (CB)^{-1}p \operatorname{sgn}(s(t)) \tag{20}$$

The proposed controller guarantees the hitting condition is met and the proof can be found in [9][10]. Notice that the signum function, $\operatorname{sgn}(t)$, is a switching function that cause chattering. However, in this work the main objective is vibration elimination and chattering elimination is not the main focus. Thus the function is maintained in the controller.

4. Simulations and results

The simulation of the AMB system and proposed controller is performed by using the following parameters shown in Table 1 [3][4].

Table 1: Parameters for Active Magnetic Bearing

Parameter	Symbol	Value
Mass of rotor (kg)	m	13.9
Moment of inertia about X ($kg.m^2$)	J_x	0.0134
Moment of inertia about Y ($kg.m^2$)	J_y	0.232
Half the longitudinal length (m)	l	0.13
Steady state gap length (mm)	G_o	0.55
Rotational speed (rad/sec)	p	0- $\pi 50$
Coil resistance (ohm)	R	10.7
Coil inductance (H)	L	0.285
Radial eccentricity (N/m)	α	0
Steady state attractive force for left and right (N)	$F_{l1,r1}$ $F_{l1-l4,r2-r4}$	90.9 22
Steady state current for left and right coils (A)	$I_{l1,r1}$ $I_{l1-l4,r2-r4}$	0.63 0.31

For the selection of the controller gain K , pole placement method is used in which the poles is selected to reside on the left-hand plane to ensure system stability. Thus for this case, two set of 3 poles and a set of 6 poles are selected for the above 2 SISO and 1 MIMO system. The calculated K values for each subsystems with their respective poles selections are as follows:

- 1st subsystem: $K_1 = 10^7 [1.2016 \ 0.0001 \ 0.0008]$, for poles $\lambda_1 = \{-7500, -10555, -11100\}$,
- 2nd subsystem: $K_2 = 10^7 [-1.0804 \ -0.0001 \ 0.0006]$, for poles $\lambda_2 = \{-5900, -6555, -7100\}$,
- 3rd subsystem:

$$K_3 = 10^6 \begin{bmatrix} 2.1721 & 0.0001 & 0.0085 & 0.1102 & 0.0 & 0.0006 \\ 0.4246 & 0.0 & 0.0016 & 1.3757 & 0.0001 & 0.0086 \end{bmatrix}$$

for poles

$$\lambda_3 = \{-6500 -6900 -8100 -9900 -11900 -16700\}.$$

The same gain K for all the subsystems are used for feedback control (pole placement) used to compare the performance of the designed controller to overcome the vibration. Next for the selection of the controller matrix C , the following values are chosen:

1st subsystem: $C_1 = [3.8493 \ 0.0002 \ 0.0019],$

2nd subsystem: $C_2 = [9.8522 \ 0.0002 \ -0.0045],$

3rd subsystem:

$$C_3 = 10^4 \begin{bmatrix} 3.7370 & 0.0001 & 0.0115 & 0.3608 & 0.0 & 0.002 \\ 4.2208 & 0.0001 & 0.0130 & 0.4075 & 0.0 & 0.0023 \end{bmatrix}$$

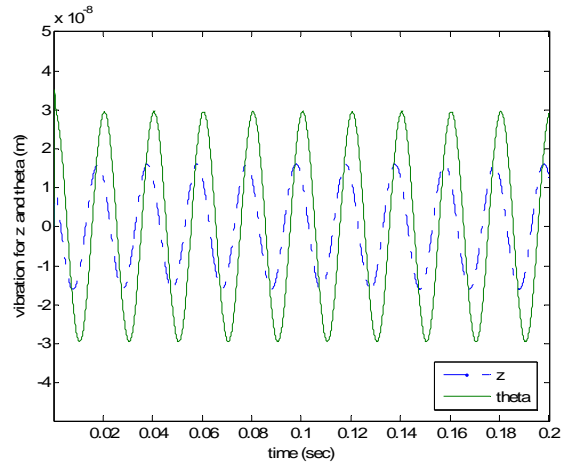


Fig. 4: The displacement for z -axis and θ motion

In selecting the values for matrix C , a simple optimization algorithm is used in which a set of simultaneous equations are solved and these values are used to access the system performance. The process is repeated until the performance of the system is satisfied.

Vibration magnitude below 0.5 micron is considered an acceptable performance for most AMB system. Figure 2 above illustrates the the gap deviation, g_{13}' , which is the magnitude of vibration on the left horizontal of the system. The amplitude of the vibration is only about 0.016 micron which is very low and considered zero relative to the total gap. Comparing to pole placement method, the amplitude of vibration is 0.35 micron which is close to upper side of acceptable range, the proposed controller is able to suppress the vibration about 22 times further. Figure 3 shows the gap deviation, g_{11}' , on the left vertical of the system. The same performance can be noticed in which the magnitude of vibration under the robust controller is extremely low and equivalent to zero. Figure 4 then shows the component of vibration in the z -axis motion and θ motion. These two components contribute to the vibration of rotor in vertical movement. The figure shows the magnitude in both motions under the proposed controller. In this case the vibration in z -axis motion is a little larger than the vibration in θ motion which suggest that the C parameter for the SISO subsystem of z - motion can further be tuned. Figure 5 illustrates the sinusoidal magnetic force produced by the coil in order to compensate the vibration term present in the system.

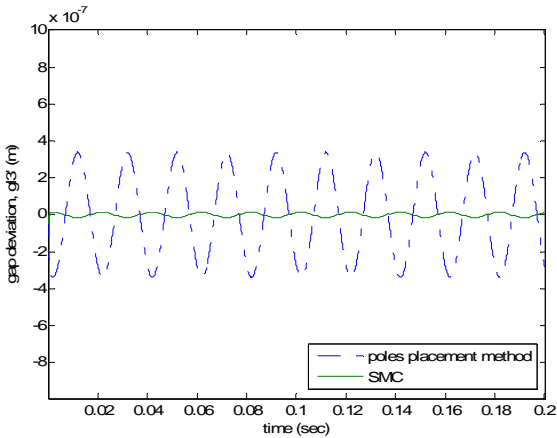


Fig. 2: Comparison of gap deviation for g_{13}' between robust controller and pole place method.

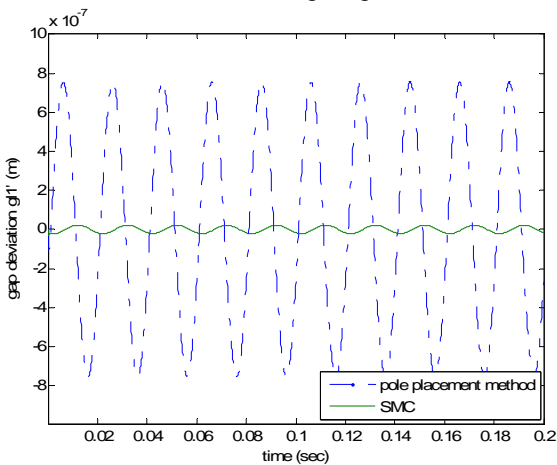


Fig. 3: Comparison of gap deviation for g_{11}' between robust controller and pole place method.

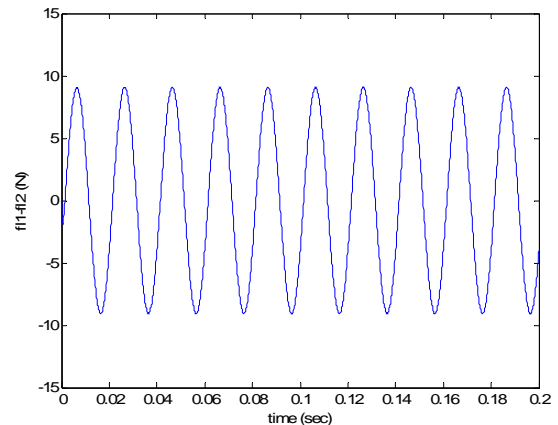


Fig. 5: Magnetic force acting on the magnetic bearing.

5. Conclusion

This paper concerns with design of full order sliding mode control method that can eliminate the rotor vibration due to mass imbalance. After deriving the dynamic model of the system and presented in state-space representation, it is shown that this vibration term also suffers mismatched condition in which it is not in phase with the input voltage. Based on proportional integral sliding mode design, the robust controller proposed is able to suppress the vibration and solve the mismatched problem. Simulation results show that the vibration is kept to a very minimum range and fulfil the design requirement. However, tuning of the controller parameter is quite time consuming and a more intelligent method will improve the overall design. This will be the next step of this work.

REFERENCES

- [1] Lee J.H.; Allaire P. E.; Tao G.; Decker J.; and Zhang X. 2003. Experimental Study of Sliding Mode Control for a Benchmark Magnetic Bearing System and Artificial Heart Pump Suspension. *IEEE Trans. on Contr. Sys. Tech.*, vol. 11, no. 1.
- [2] Matsumura F.; and Yoshimoto T. 1986. System Modeling and Control Design of a Horizontal-shaft Magnetic Bearing System. *IEEE Trans. on Magnetics.*, vol. MAG-22, no. 3.
- [3] Matsumura F.; Fujita M.; and Okawa K. 1990. Modeling And Control of Magnetic Bearing Systems Achieving A Rotation Around the Axis of Inertia. In proceedings of the 2nd International Symposium on Magnetic Bearings, 273-280. Tokyo, Japan.
- [4] Mohamed A. M.; Matsumura F.; Namerikawa T.; and Lee J. H. 1997. Q-Parameterization Control of Vibrations In A Variable Speed Magnetic Bearing. In *Proceedings Of IEEE International Conference on Control Applications*, 540-545. Hartford, CT.
- [5] Mohamed A. M.; Hassan I. M.; Hashem A. M. K. 1999. Elimination of Imbalance Vibrations in Magnetic bearing System Using Discrete-Time Gain-Scheduled Q-Parameterization Controllers. In *Proceedings Of IEEE Conf. Int. Conf. on Contr Application*, 737-741. Kohala Coast-Island of Hawai'i, Hawai'i.
- [6] Herzog R. ; Buhler P. ; Gahler C. ; and Larsonneur R. 1996. Unbalance Compensation Using Generalized Notch Filters in the Multivariable Feedback of Magnetic Bearings. *IEEE Transactions on Control Systems Technology*. Vol. 4, No. 5.
- [7] Mohamed A. M.; and Emad F. P. 1993. Comparison Between Current and Flux Control in Magnetic Bearing Systems. In *Proceedings. of the American Control. Conference*.
- [8] Yan J.J.; Tsai J.S.-H; and Kung F. -C. 1997. Robust Decentralized Stabilization of Large-Scale Delay Systems Via Sliding Mode Control. *ASME Journal of Dynamic Systems, Measurement and Control*. 307-312.
- [9] Sam Y. M.; and Osman J. H. S. 2003. Active Suspension Control: Performance Comparison using Proportional Integral Sliding Mode and Linear Quadratic Regulator Methods. In *Proceedings in Control Application Conference*. 274-278.
- [10] Cristi R. ; Papoulias F. A. ; and Healey A. J. 1990. Adaptive Sliding Mode Control of Autonomous Underwater Vehicles in the Dive Plane. *IEEE Journal of Oceanic Engineering*. Vol 15, No. 3.

Analytical and Empirical Study of Particle Swarm Optimization with a Sigmoid Decreasing Inertia Weight

Andi Adriansyah¹, Shamsudin H. M. Amin²

¹Department of Electrical, Faculty of Industrial Engineering,
Universitas Mercu Buana, Jl.Raya Meruya Selatan, Kembangan, Jakarta, 11650, Indonesia
Phone: +6221-5840816, Fax : +6221-5840813, E-mail: aandi2@siswa.utm.my

²Centre of Artificial Intelligence and Robotics (CAIRO), Faculty of Electrical Engineering,
Universiti Teknologi Malaysia, Skudai, 81310, Johor Bahru, Malaysia
Phone: +607-5535319, Fax: +607-5566272, E-mail: sham@fke.utm.my

Abstract

The particle swarm optimization (PSO) is an algorithm for finding optimal regions of complex search space through interaction of individuals in a population of particles. Search is conducted by moving particles in the space. Some methods area attempted to improve performance of PSO since is founded, including linearly decreasing inertia weight. The present paper proposes a new variation of PSO model where inertia weight is sigmoid decreasing, called as Sigmoid Decreasing Inertia Weight. Performances of the PSO with a SDIW are studied analytically and empirically. The exploration–exploitation tradeoff is discussed and illustrated, as well. Four different benchmark functions with asymmetric initial range settings are selected as testing functions. The experimental results illustrate the advantage of SDIW that may improve PSO performance significantly.

Keywords:

Particle Swarm Optimization; Inertia Weight; Sigmoid Decreasing Inertia Weight

1. Introduction

The difficulties associated with using mathematical optimization on large-scale engineering problem have contributed to the development of alternative solutions. To overcome these problems, researchers have proposed evolutionary-based algorithms for searching near-optimum solutions to problems. Evolutionary algorithms are stochastic search methods that mimic the metaphor of natural biological evolution and/or the social behavior or species. To mimic the efficient behavior of these species, various researchers have developed computational systems that seek fast and robust solutions to complex optimization problems. Particle Swarm Optimization (PSO) is one of evolutionary computation technique developed by Kennedy and Eberhart in 1995 [1, 2]. The method finds the optimal solution by simulating such social behavior of groups as fish schooling or bird flocking. A group can achieve the objective effectively by using the common information of every particle, and the information owned by the particle itself.

However, the PSO algorithm includes some tuning parameters that greatly influence the algorithm performance, known as the exploration-exploitation tradeoff. Balancing between exploration and exploitation searching process will improve PSO performance. A number of methods have been provided to get to the bottom of the problem. Early experience with PSO was proposed by Shi and Eberhart that introduced inertia weight and maximal velocity which

tuned based on trial and error [3]. Suitable selection of the inertia weight provides a balance between global and local searching. Afterward, they presented a new concept about inertia weight [4]. In this concept, they attempted to get better of PSO performance by linearly decreasing inertia weight (LDIW). They also tried to overcome the problem by changing inertia weight adaptively based on Fuzzy System [5] and randomly [6]. Furthermore, recent work done by Clerc [7] indicates that use of a constriction factor may be necessary to insure convergence of the PSO. In constriction factor, inertia weight adjusted concurrently with another PSO parameters. In contrast, Zheng et. al., investigated increasing inertia weight in their research [8]. According to them, a PSO with increasing inertia weight outperforms the one with decreasing inertia weight. Though, the results still not satisfied.

This paper presents an approach to overcome exploration-exploitation tradeoff problem. A new nonlinear function modulated inertia weight adaptation with time proposed for improved performance of PSO algorithm. Instead of linearly decreasing inertia weight, the schema attempted to decrease inertia weight by means of sigmoid function. In this work, some analytical and empirical studies are investigated. In section 2, philosophy and procedure of original PSO are explained. Some analysis also presented in this section. In section 3, a new PSO model with a sigmoid decreasing inertia weight (SDIW) is suggested. To prove the validity of such method, several standard benchmark functions are tested in Section 4. The

empirical data resulted will be emphasized and discussed in Section 5. Finally Section 6 concludes this paper.

2. PSO Algorithm and Analysis

2.1. Philosophy and Procedure of Standard PSO

PSO is one of the artificial life or multiple particles' type techniques designed and developed by Kennedy and Eberhart [1, 2]. The concept of original PSO can be described as follows: each potential solution, called *particle*, knows its best value so far (*pbest*) and its position. Moreover, each particle knows the best value in the group (*gbest*) among the *pbest*. All of the best values are based on fitness function ($F(.)$) for each problem to be solved. Each particle tries to modify its position using the current velocity and its position. The velocity of each particle can be calculated using the following Equation:

$$v_i^{k+1} = v_i^k + c_1 \text{rand}() (pbest_i^k) + c_2 \text{rand}() (gbest_i^k) \quad (1)$$

where v_i^k , v_i^{k+1} , and s_i^k , are velocity vector, modified velocity and positioning vector of particle i at generation k , respectively. Then, *pbest* and *gbest* are best position found by particle i and best position found by particle group. Finally, c_1 and c_2 are cognitive and social coefficients, respectively, that influence particles velocity. Afterward, the current position of a particle is calculated by the following Equation:

$$s_i^{k+1} = s_i^k + v_i^{k+1} \quad (2)$$

Updating process of velocity and position of each particle is depicted in Fig. 1.

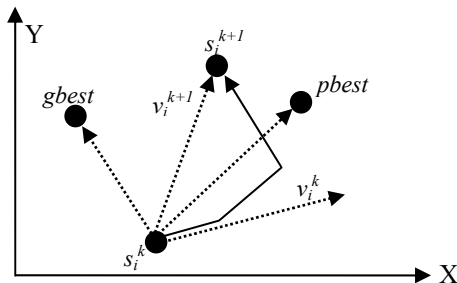


Figure 1. The velocity and position updates in PSO

An algorithm to find the best positioning vector of PSO using n particles can be summarized as follows:

1. Initial positioning vector $S[n]$ and velocity vector $V[n]$ are generated by using random values, where $s_i = [s_i^1, s_i^2, \dots, s_i^n]$ and $v_i = [v_i^1, v_i^2, \dots, v_i^n]$.
2. Velocity vector v_i^{k+1} of particle i is calculated by using Equation (1).
3. New positioning vector s_i^{k+1} of particle i is calculated by using Equation (2).
4. If $F(s_i^k)$ is better than the $F(pbest_i)$, the positioning vector s_i^k is set to *pbest*. If $F(pbest_i)$ is better than $F(gbest)$, the positioning vector *gbest* is set to *pbest*.
5. If the generation reaches to the pre-determined one, process will stop. Otherwise, will go to step 2.

In order to get better control exploration and exploitation of particles searched, the concept of inertia weight, w , is developed [3, 4]. Introducing inertia weight concept, Equation (1) and Equation (2) can be written as:

$$v_i^{k+1} = w_i v_i^k + c_1 \text{rand}() (pbest_i^k) + c_2 \text{rand}() (gbest_i^k) \quad (3)$$

$$s_i^{k+1} = s_i^k + v_i^{k+1} \quad (4)$$

Almost previous researches on PSO system have provided empirical results and informal analyses. This paper presents formal analysis of traditional simple particle systems, crucial to understand the dynamics of how particle behavior depends on parameters for making the right choice of parameter values.

2.2. Particle Trajectory Analysis

It appears in Equation. (3) and (4) that each dimension is updated independently from the others. The only link between the dimensions of the problem space is introduced via the objective function, i.e., through the locations of the best positions found so far *pbest* and *gbest*. In order to understand the behavior of a complex system, it often helps to begin by examining a simpler version of it. Thus, without loss of generality, the algorithm description can be reduced to the one-dimensional case:

$$v_{k+1} = wv_k + c_1 \text{rand}() (pbest - s_k) + c_2 \text{rand}() (gbest - s_k) \quad (5)$$

$$s_{k+1} = s_k + v_{k+1} \quad (6)$$

For the theoretical analysis of the PSO, the deterministic version will be considered. The deterministic version is obtained by setting the random numbers to their expected values:

$$\text{rand}() = \text{rand}() = 1/2 \quad (7)$$

Thus, Equation (7) can be simplified using the notation:

$$c = \frac{c_1 + c_2}{2}, \quad (8)$$

$$p = \frac{c_1}{c_1 + c_2} pbest + \frac{c_2}{c_1 + c_2} gbest \quad (9)$$

Using this notation, the deterministic PSO algorithm can be expressed as:

$$v_{k+1} = wv_k + c(p - s_k) \quad (10)$$

$$s_{k+1} = s_k + v_{k+1} \quad (11)$$

The algorithm described by Equation (10) and (11) contains two tuning parameters, w and c , that are truly influence for PSO performance.

In order to analyze dynamic system of PSO, Equation (10) and (11) can be combined and written in compact matrix form as follows:

$$z_{k+1} = Az_k + Bp \quad \text{with}$$

$$z_k = \begin{bmatrix} v_k \\ s_k \end{bmatrix}, \quad A = \begin{bmatrix} w & c \\ w & I + c \end{bmatrix}, \quad B = \begin{bmatrix} b \\ b \end{bmatrix} \quad (12)$$

In the context of dynamic system theory, z_k is the particle state made up of its current position and velocity, A is the dynamic matrix whose properties determine the time behavior of the particle, p is the external input used to drive the particle towards a specified position and B is the input matrix that gives the effect of the external input on the particle state.

Standard results from dynamic system theory say that the time behavior of the particle depends on the eigenvalues of the dynamic matrix A . The eigenvalues λ_1 and λ_2 (either real or complex) are the solutions of characteristic polynomial equation:

$$\lambda^2 + (c - w - 1)\lambda + w = 0 \tag{13}$$

The necessary and sufficient condition to be stable and will converge is that both eigenvalues of the matrix A have magnitude less than 1. The analysis of the roots of Equation (13) leads to the following set conditions:

$$w < 1, \quad c > 0, \quad \text{and } 2w - c + 2 > 0 \tag{14}$$

The convergence domain in the (w,c) plane is the triangle shown in Fig. (2). For any initial position and velocity, the particle will converge to its equilibrium position, p , as in Equation. (10) if and only if the algorithm parameters are selected inside this triangle.

Before convergence, the particle exhibits harmonic oscillation around the equilibrium point when the eigenvalues of the matrix A , which are also the roots of Equation (14), are complex. This equivalent to:

$$w^2 + c^2 - 2wc - 2w - 2c + 1 < 0 \tag{15}$$

The corresponding domain in the (w,c) plane is elliptical function as depicted in Fig. (2).

The particle may also exhibit zigzagging behavior around the equilibrium point when at least one of the eigenvalues of the matrix A , whether real or complex, has negative real part. This is equivalent to:

$$w < 0 \quad \text{or} \quad w - c + 1 < 0 \tag{16}$$

The corresponding domain in the (w,c) plane is drawn in Fig. (2). Zigzagging may be combined with harmonics oscillation.

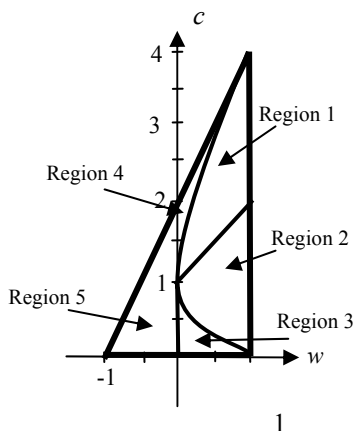


Figure 2. Domain of dynamic behavior in the (w,c) parameter space

2.3. Convergence Analyses

There are several interesting regions for the analyses based on integration of the three figures aforementioned before, such as shown in Fig. (2). In Region 1, the eigenvalues are complex number with real positive number, $\text{Re } \lambda_1$ and $\text{Re } \lambda_2 > 0$. In this region, the particle exhibits harmonic oscillation around the equilibrium point before convergence. In Region 2, the eigenvalues are complex number as well, but with negative real number, $\text{Re } \lambda_1$ and $\text{Re } \lambda_2 < 0$. In this region, the particle exhibits combination of harmonic oscillation and zigzagging around the equilibrium point before convergence. Furthermore, in Region 3, the eigenvalues are positive real number, λ_1 and $\lambda_2 > 0$, where the particle in this region directly convergence to equilibrium point without harmonic oscillation and zigzagging. In Region 4, the eigenvalues are negative real number, λ_1 and $\lambda_2 < 0$. The particle exhibits symmetric zigzagging around the equilibrium point before convergence without harmonic oscillation. Finally, in Region 5, the eigenvalues are positive and negative real number, $\lambda_1 > 0$ and $\lambda_2 < 0$. In this region, the particle exhibits asymmetric zigzagging around the equilibrium point before convergence without harmonic oscillation. Outside of these regions, the particle will diverge. Particle trajectories of these regions are depicted in Fig. (3).

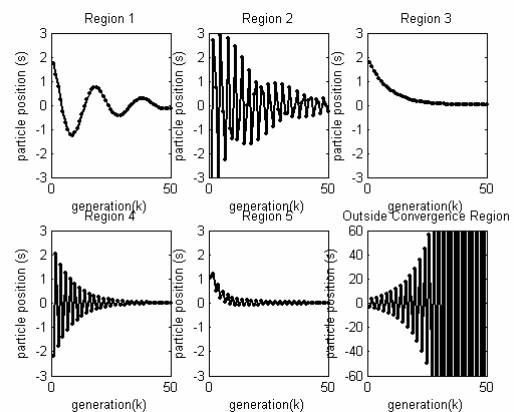


Figure 3. Particle Trajectories and Regions

However, beside determine the shape of particle trajectories, parameter couples agree on speed of convergence. As a general rule, the value of w and c which close to the center of the stability triangle induce quick convergence, while the values close to its borders require many generations to converge, as illustrated in Fig. (4).

According to convergence analysis aforementioned before, a large inertia weight has slow convergence to facilitate a global search, while, a small inertia weight has quick convergence to facilitate a local search. Therefore, by decreasing the inertia weight, w , from a relatively large value to a small one with associated cognitive and social coefficients, c , through the course of the PSO run, the PSO tends to have more global search ability at the beginning of

the run and have more local search ability near the end of the run.

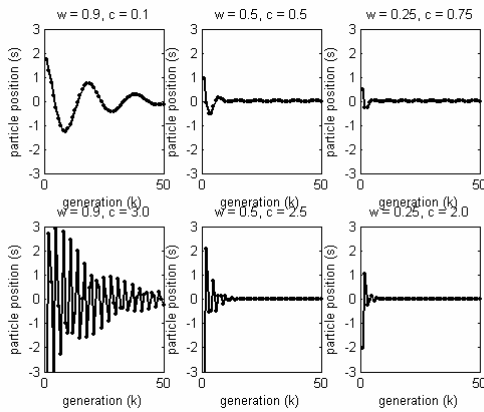


Figure 4. Particle Trajectories and Regions

However, how to decrease inertia weight effectively to achieve good balancing of global and local search is not an easy task. Sigmoid decreasing inertia weight (SDIW) is implemented in this paper to find a better compromise of exploitation-exploration trade-off.

3. PSO with Sigmoid Decreasing Inertia Weight

The present paper proposes a new nonlinear function modulated inertia weight adaptation with time for improved performance of PSO algorithm. Instead of linearly decreasing of inertia weight, the schema attempted to decrease inertia weight by means of sigmoid function, as shown in Fig. 5.

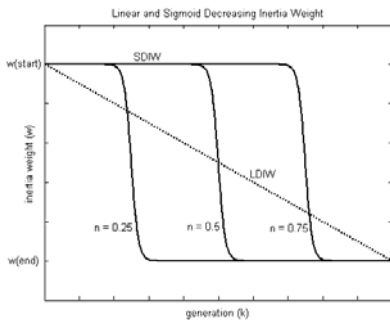


Fig. 5. Sigmoid and Linearly Decreasing Inertia Weight

The proposed function of sigmoid is given as:

$$w_k = \frac{w_{start} (w_{start} - w_{end})}{(1 + e^{p*(k/n - 1)})} \tag{17}$$

when

$$p = 10^{(\log(\text{gen}) - 2)} \tag{18}$$

where w_k is inertia weight at k , w_{start} and w_{end} are inertia weight at the start and the final inertia weight at the end of a given run, respectively. Furthermore, p is the constant to adjust sharpness of the function, gen is the maximum

number of generations to run and n is the constant to set partition of sigmoid function.

In sigmoid, a large inertia weight is maintained in first part of PSO process to assure a global search. Afterwards, a small inertia weight is retained to facilitate a local search in final part of PSO process. There is very short inertia weight gradation between large and small one. This method will provide a balance between global and local searching to give the PSO a superior performance.

In order to further illustrate the effect of this SDIW, some experiments are set. The results are shown and discussed in next sections.

4. Experimental Setting

To illustrate the behavior of the proposed method, four non-linear benchmark functions are used here. The first function is the Sphere function described by Equation (19):

$$f_0(x) = \sum_{i=1}^n x_i^2 \tag{19}$$

where $x = [x_1, x_2, \dots, x_n]$ is an n -dimensional real-valued vector. The second function is the Rosenbrock function given as:

$$f_1(x) = \sum_{i=1}^n (100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2) \tag{20}$$

The third function is the generalized Rastrigin function formulated as:

$$f_2(x) = \sum_{i=1}^n (x_i^2 - 10 \cos(2\pi x_i) + 10) \tag{21}$$

The fourth function is the generalized Griewank function shown as:

$$f_3(x) = \frac{1}{400} \sum_{i=1}^n x_i^2 \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1 \tag{22}$$

For the purpose of evaluation, the asymmetric initialization method is adopted her for the population initialization. Table 1 lists the initialization ranges of the four functions:

Table 1. Asymmetric Initialization Range

Function	Asymmetric Initialization Range
f_0	$(50, 100)^n$
f_1	$(15, 30)^n$
f_2	$(2.56, 5.12)^n$
f_3	$(300, 600)^n$

For each function, three different dimension sizes are tested. They are dimension sizes: 10, 20 and 30. The maximum number of generations is set as 1000, 1500 and 2000 corresponding to the dimensions 10, 20 and 30, respectively. In order to investigate whether the PSO algorithm scales well or not, different population sizes are used for each function with different dimensions. They are population sizes: 20, 40 and 80. A sigmoid decreasing inertia weight is used which starts at 0.9 and ends at 0.4, which $c_1 = 2$ and $c_2 = 2$. With aim to find the best partition of sigmoid function, different sigmoid constants, n , are used, that are: 0.25, 0.5 and 0.75.

5. Experimental Results

As the main objective of searching methods was to achieve faster speed convergence and better solution accuracy, the experiments results will be shown in table and graphs. Linear Decreasing Inertia Weight (LDIW) used as comparison to proposed method.

Figure 6 shows the results for the sphere function with two different population sizes, respectively. Table 2 lists the mean fitness values of the best particle found for the 30 runs for each function. It is easy to see for the sphere function, that PSO can find the more optima values vary fast and PSO algorithm also scales very well, especially for n equal to 0.25.

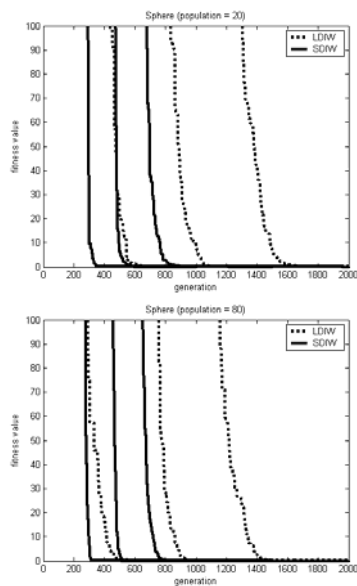


Figure 6. Curve of Sphere function

Table 2. The mean fitness values for the Sphere function

Dim	LDIW	SDIW		
		n = 0.25	n = 0.5	n = 0.75
Pop = 20				
10	2.6897e-09	5.8519e-35	7.2656e-22	1.3174e-10
20	2.9122e-07	3.5893e-21	9.0725e-08	3.1900e-04
30	5.5721e-05	2.6329e-15	3.7966e-08	3.0000e-03
Pop = 40				
10	4.3735e-18	1.6285e-42	6.8196e-27	1.0791e-11
20	6.2059e-10	1.6278e-28	4.9444e-17	1.3624e-06
30	1.0369e-06	4.6623e-21	1.3846e-12	2.9829e-04
Pop = 80				
10	8.7638e-19	7.3267e-49	1.3887e-31	1.1554e-12
20	5.7939e-12	6.9927e-32	1.8691e-19	1.4061e-09
30	2.1866e-08	5.6160e-25	2.3648e-15	2.6464e-06

Figure 7 shows the results for the Rosenbrock function with two different population sizes, respectively. Figure 8 and Figure 9 show the results for the Rastrigin and Griewank functions with two different population sizes, respectively.

Table 3 to 5 list the mean fitness values of the best particle found for the 30 runs for the other three functions, respectively.

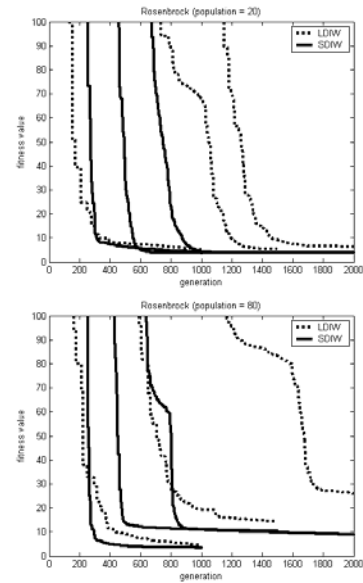


Figure 7. Curve of Rosenbrock function

Table 3. The mean fitness values for the Rosenbrock function

Dim	LDIW	SDIW		
		n = 0.25	n = 0.5	n = 0.75
Pop = 20				
10	5.0990	3.8001	4.4028	6.2818
20	5.3858	3.9866	4.0420	5.3584
30	6.2131	3.9866	4.0801	5.8690
Pop = 40				
10	4.8889	3.3509	4.0093	5.4973
20	4.3776	4.1118	3.9998	8.6896
30	5.5788	4.4133	4.2580	10.5898
Pop = 80				
10	4.5707	3.3854	3.9582	5.1686
20	14.3539	10.4202	6.5787	14.5359
30	18.3061	8.8637	4.0649	4.1416

By looking at the shape of the curves in all the figures, it is easy to see the proposed PSO converges quickly under all cases and will slow its convergence speed down when reaching the optima values. Nevertheless, the results shown illustrate that by using a SDIW, the performance of PSO can be improved greatly and have better results than LDIW. From the figures, it is also clear that the PSO with different population sizes has almost the similar performance.

6. Conclusion

In this paper, particle trajectories of simple PSO process have been analyzed. Significant parameters, inertia weight and cognitive and/or social constants, have been studied. The sigmoid function to improve PSO performance has been described. The performance of the PSO algorithm with Sigmoid Decreasing Inertia Weight has been investigated and extensively compared with Linear Decreasing Inertia Weight by experimental studies of four

nonlinear functions. Experiments results shows that propose method has more robust property with the available PSO method, that is, not sensitive to population size in all four functions test. Moreover, the new PSO method greatly improves the accuracy and convergence speed of search, especially for sigmoid function with n equal 0.25.

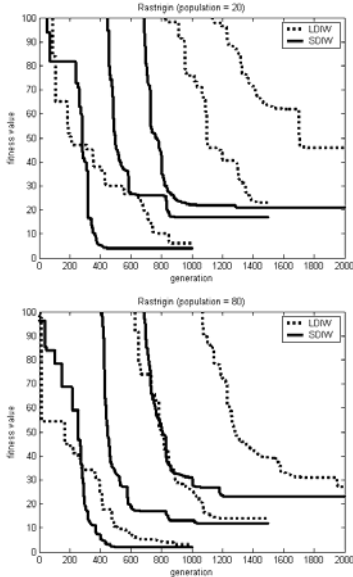


Figure 8. Curve of Rastrigin function

Table 4. The mean fitness values for the Rastrigin function

Dim	LDIW	SDIW		
		n = 0.25	n = 0.5	n = 0.75
Pop = 20				
10	5.9698	3.9798	4.9748	7.9597
20	22.8917	16.9143	19.8992	23.9236
30	45.7682	20.8941	40.7933	49.4858
Pop = 40				
10	3.9798	2.9489	2.9849	4.9760
20	15.9194	11.8387	14.9244	17.2506
30	40.7939	31.8387	38.8033	46.7630
Pop = 80				
10	2.9849	1.9899	3.9798	4.0131
20	13.9294	11.9395	12.9345	15.9195
30	27.4395	22.8840	27.8588	28.9393

Table 5. The mean fitness values for the Griewank function

Dim	LDIW	SDIW		
		n = 0.25	n = 0.5	n = 0.75
Pop = 20				
10	0.0984	0.0836	0.0763	0.0960
20	0.0713	0.0662	0.0636	0.0320
30	0.0742	0.6364	0.0271	0.0999
Pop = 40				
10	0.0640	0.0787	0.0684	0.0935
20	0.0835	0.0711	0.0614	0.1250
30	0.0737	0.0246	0.0197	0.0866
Pop = 80				
10	0.0813	0.0713	0.0615	0.0866
20	0.0492	0.0246	0.0123	0.0710
30	0.0172	0.0148	0.0123	0.0197

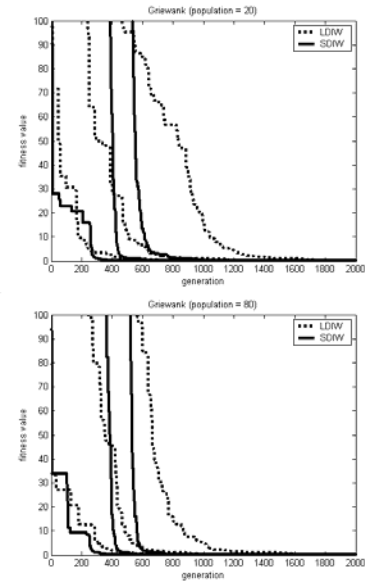


Figure 9. Curve of Griewank function

References

- [1] Kennedy, J. and Eberhart, R.C. 1995. Particle swarm optimization, In *Proceeding of IEEE International Conference on Neural Networks, IV*, 1942-1948. Piscataway, NJ: IEEE Service Center.
- [2] Eberhart, R. C., and Kennedy, J. 1995. A new optimizer using particle swarm theory, In *Proceeding of the Sixth International Symposium on Micro Machine and Human Science*, 39-43. Nagoya, Japan, Piscataway, NJ: IEEE Service Center.
- [3] Shi, Y., Eberhart, R. C. 1998. Parameter selection in particle swarm optimization. In *Evolutionary Programming VII*, 591-600. New York: Springer-Verlag.
- [4] Shi, Y., and Eberhart, R.C. 1998. A modified particle swarm optimizer. In *Proceeding of the IEEE International Conference on Evolutionary Computation*, 69-73. NJ: IEEE Press, Piscataway.
- [5] Shi, Y., and Eberhart, R. C., 2001. Fuzzy Adaptive Particle Swarm Optimization, In *Proceeding Congress on Evolutionary Computation 2001*, 101-106. Seoul, Korea, Piscataway, NJ: IEEE Service Centre.
- [6] Eberhart, R. C., and Shi, Y. 2001. Tracking and optimizing dynamic systems with particle swarms, In *Proceeding Congress on Evolutionary Computation 2001*, Seoul, Korea, Piscataway, NJ: IEEE Service Centre.
- [7] Clerc, M., and Kennedy, J., 2002. The Particle Swarm: Explosion, Stability, and Convergence in a Multi-Dimensional Complex Space, *IEEE Transaction on Evolutionary Computation* 6(1):58-73.
- [8] Zheng, Y, et. al., 2003. Empirical Study of Particle Swarm Optimizer with an Increasing Inertia Weight, In *Proceeding of the IEEE Congress on Evolutionary Computation 2003*, Vol.1, 221-226.

A True Random Number Generator for Crypto Embedded Systems

Norashikin M. Thamrin¹, Illiasaak Ahmad¹, Mohamed Khalil Hani¹

¹VLSI-ECAD Research Laboratory, Microelectronic and Computer Engineering Department (MiCE)
Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia.
Tel: +607-5535223, Fax: +607-5566272, E-mail: norashikin.mthamrin@gmail.com,
ilyasak_ahmad@yahoo.com, khalil@fke.utm.my

Abstract

In this paper, we design a true random number generator (TRNG) in hardware which is targeted for FPGA-based crypto embedded systems. All crypto protocols require the generation and use of secret values that must be unknown to attackers. Random number generators (RNG) are required to generate public/private key pairs for asymmetric algorithm such as RSA and symmetric algorithm such as AES. Since security protocols rely on the unpredictability of the keys they use, RNGs for crypto applications must meet stringent requirements. The most important in cryptography is that attackers must not be able to make any useful predictions about the RNG outputs. The TRNG employs internal analog phase-locked loop (PLL) circuitry to generate a noise which is useful in producing random output. In contrast with traditionally used free running oscillators, it uses a novel method of random noise extraction based on two rationally related synthesized clock signals. The digital design is extremely compact and can be implemented on any advance FPGA device equipped with analog PLL. With the help of this TRNG, the cryptographic implementations such as key generation, authentication protocols, digital signature schemes and even in some encryption algorithm can be secure enough from being abused by malicious act.

Keywords

True Random Number Generator, Synthesized Clock, PLL, Embedded system, Cryptography

1. Introduction

The issue of random number generator is becoming crucial for implementation of cryptographic systems in Field Programmable Gates Array (FPGAs). Key generation, authentication protocols, zero-knowledge protocols, and block padding challenges are some of the cryptographic objects where a string of unpredictable bits or random numbers are required. In all these applications, security greatly depends on the quality of the source of the randomness. The quality of generated numbers is proved by statistical tests. In addition to good statistical properties of the obtained numbers, the output of the generator used in cryptography must be unpredictable. For this reason, pseudo-random number generators, which are deduced through algorithmic process, are not suitable for cryptographic applications.

TRNGs can be produced using any non-deterministic process. This means that special attention should be paid to avoid weaknesses that help the attacker to break a system. The design is aimed to be implemented in embedded crypto system. Digital circuits of modern Field Programmable Logic Devices (FPLDs) include only limited source of randomness, e.g. metastability, frequency of free-running oscillators, clock jitter, etc. Usually, reliable and fast TRNG based on metastability is hard to implement and not secure enough in cryptographic applications. Free running oscillators are typically used in known FPGA based TRNGs. In principle, TRNGs based on free running oscillator and intrinsic jitter contained in digital circuits can be used without any additional FPLD resources.

In this paper, we present a TRNG design of a novel method of randomness extraction based on two rationally related synthesized stable clock signals. It was shown that it is well suited for modern FPLDs with internal analog Phase-Locked Loop (PLL) circuitry (e.g. Apex, Cyclone or Stratix FPLDs from Altera). We use the large flexibility of PLLs embedded in Stratix FPLDs to demonstrate the relationship between PLL and TRNG configurations, the quality of the output random bit-stream, and the speed of the generator. Although the TRNG was developed for the Altera Stratix family of devices, the principle can be easily employed in other digital devices containing analog PLLs.

2. The PLL-Synthesized Clock Jitter

2.1. Analog PLLs embedded in digital circuits

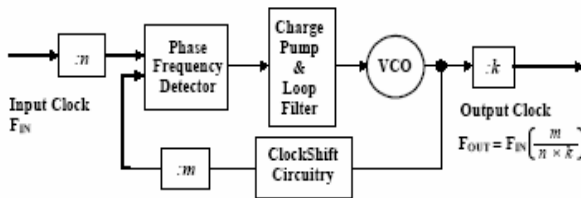
A digital VLSI circuit uses advanced clock generation and distribution circuitry based on embedded analog PLLs. A simplified block diagram of one analog based PLL block available in advanced digital circuits is depicted in Figure 1. Each PLL block can provide at least one synthesized clock signal with frequency F_{OUT}

$$F_{OUT} = F_{IN} [m / (n \cdot k)] = F_{IN} [K_M / K_D] \quad (1)$$

where F_{IN} is the frequency of the external input clock source. Reference-, feedback- and post-divider values n , m and k can vary from one to several hundreds in FPGAs.

The Altera Stratix devices include two types of PLLs[7]:

- **Fast PLL (FPLL):** Stratix devices include up to 8 FPLLs. The FPLLs offer general-purpose clock management with multiplication and phase shifting. The multiplication is implied in comparison to EPLL and uses only $m=k$ scaling factors with a range from 1 to 32. Input frequency can vary in dependency on m (for speed grade -5) from 15 to 717 MHz, output frequency from 9.4 to 420 MHz, and the frequency of the Voltage Control Oscillator (VCO) from 300 to 1000 MHz.
- **Enhanced PLL (EPLL):** Compared to FPLL, EPLLs have additional configurable features like external feedback, configurable bandwidth, run-time reconfiguration, etc. They also have an enhanced range of parameters. The input frequency can vary for a speed grade -5 device from 3 to 684 MHz, output frequency from 9.4 to 420 MHz and the frequency of the VCO from 300 to 800 MHz. Reference-, feedback- and post-divider values n , m and k can vary from 1 to 512



(1024 for k) with a 50% duty cycle.

Figure 1 Simplified block diagram of an embedded PLL circuitry.

2.2. Jitter of PLL synthesized clock signals

In analog PLLs, a noise causes the Voltage Controlled Oscillator (VCO) to fluctuate in frequency. Other frequency fluctuations are caused by variations of supply voltage, temperature and by noise environment. The internal control circuitry adjusts the VCO back to the specified frequency, but a certain part of the fluctuations caused by non-deterministic noise cannot be compensated for and is seen as a clock jitter. Clock jitter is the deviation from the ideal timing of clock transition events. As signals toggle faster and faster, the clock edge will fall outside the ideal margin. Figure 2 shows a schematic representation of clock jitter.

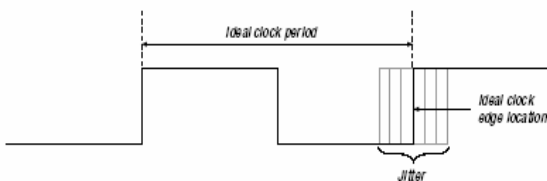


Figure 2: Clock jitter representation

2.3. Jitter generated in Stratix PLLs

The size of the intrinsic jitter depends on the quality factor Q of the VCO, on the bandwidth of the loop filter on the so-called pattern jitter introduced by the phase detector. The intrinsic jitter is often given in a peak-to-peak value or 1-sigma (RMS) value. The 1-sigma value of the jitter (σ_{jit}) depends on the technology and the configuration of the PLL and it can range up to 100ps. Since the technology of the PLL and the quality of VCO are usually defined, a user can change the output jitter directly by modification of scaling factors and filter bandwidth, but also indirectly by the design of the board (separation of the analog and digital ground, filtering of the analog power supply, etc).

Since the size of the jitter is very important in this method, previous researchers [1] have selected the Altera Development board with a Stratix EP1S25F780C5 device for jitter measurements and TRNG implementation. The jitter has been measured similarly using Agilent Infiniium DCA 86100B wide bandwidth oscilloscope. It has been found that for the FPLL with the ratio 12/7 the jitter achieves 1-sigma value of about 10 ps and for the EPLL with ratio 139/133 the 1-sigma value of the jitter is about 16 ps (see Figure 3(a) and 3(b)).

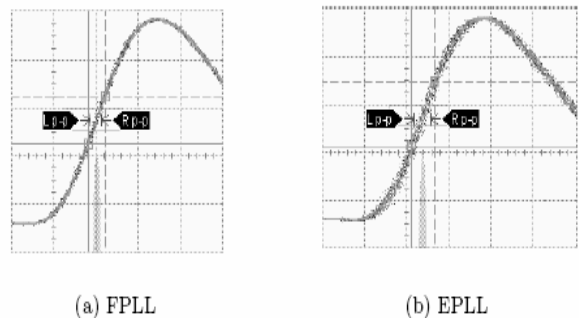


Figure 3: Jitter of the clock signal (horizontal scale: 200 ps/div)

3. TRNG Architectures Embedded in FPGAs

The basic principle behind this method is to extract the randomness from the jitter of clock signal synthesized in the embedded analog PLL. The jitter is detected by the sampling of a reference (clock) signal using rationally related (clock) signal synthesized in the on-chip analog PLL. The fundamental problem lies in the fact that the reference signal has to be sampled near the edges influenced by the jitter. Figure 4 shows the basic structure of the TRNG.

The TRNG can be designed using one or two PLLs. Since the Stratix family contains two types of PLLs, the implementation EPLL is chosen since the jitter size is bigger than FPLL. The extraction technique presented is to

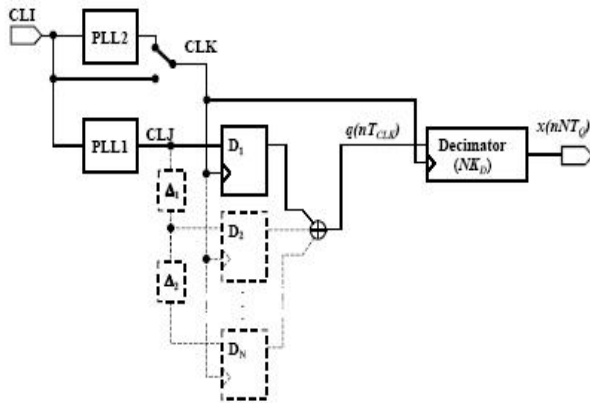


Figure 4: Basic structure of the TRNG [3]

use one clock signal to sample the value of a second clock signal on each cycle. If the two clock frequencies are slightly different, the point sampled in the second signal will advance through the second signal’s cycle. If the change is small enough it will eventually sample the second signal in the jitter zone. Thus the sampling will produce a large number of deterministic bit and at least one uncertain bit taken in the jitter zone. XORing the deterministic value and the non-deterministic bit produces a single random bit.

The proposed design as shown in Figure 5 consists of sampling circuit, serial/parallel converter, register, status register and a control unit. It uses a system clock and a synthesized PLL clock (which is identical to the system clock) as the input of the sampler circuit.

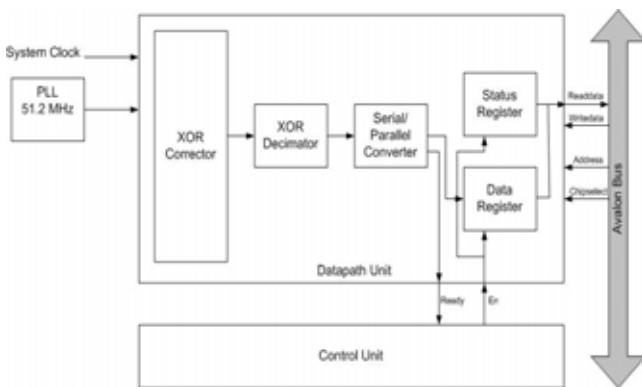


Figure 5: Overall design

The complete TRNG design is implemented on Altera Stratix EP1S40F780C5 development board. Acquired bits were transmitted to the PC through serial communication. The complete TRNG design which is including NIOS system module and serial interface controller needs up to 2,711 Logic Elements (LEs) from about 41250 LEs available in the device. The signal CLK was used as a clock signal for the control logic. The TRNG architecture was described in VHDL and implemented using Altera Quartus II development system, version 4.0. Because the jitter depends on an analog process, the real TRNG output cannot be simulated.

4. Experimental Results

The design and testing of this TRNG was done on Altera Stratix development board. Our VHDL development system is Window based. Our TRNG can produce up to 32000 bps. This paper presents results from the US National Institute of Standards and Technology (NIST) Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic applications.

The NIST test suite produces a summary report for each file of random bits it tests. Table 1 that follows is a result of running the NIST [10] suite over the set data produced by our TRNG. Each test in the NIST suite is run over a large number of sets of bits from the file to be tested. The statistic that is generated from each of these runs is called P-Value and it represents the probability of a perfect random number generator. According to NIST, if the P-Value that obtained from the test exceeds 0.001 (which is range from 0 to 1), the generator is acceptable in producing random bit. Thus, very small P-Values are bad. The documentation that accompanies the suite indicates that “If P-Value >=0.001, than the sequence can be considered random and uniformly distributed” [10]. There are 16 tests include in NIST documentation. However, the most common statistical test of TRNG is the Frequency (Monobit) test [2] [10]. In this test, our TRNG get P-Value = 0.841481 which is consider acceptable as a random number generator.

Table 1: NIST Results

NIST Test	P-Value
block-frequency	1.000000
cumulative-sums	0.499939
fft	0.771671
Frequency (monobit)	0.841481
linear-complexity	1.000000

Some of the suite tests were failed due to speed of random bit generation and limitation of bit generation. For future work, a better TRNG design can be implemented to speed up the random bit generation and can produce more random bits in one second.

5. Conclusion

This paper has described the methodology and design of PLL-based true random number generators embedded in modern FPLDs. The design uses the analog PLLs embedded in Altera Stratix FPGA family. The high quality of TRNG output was confirmed by applying special statistical tests.

The proposed solution is very cheap, uses few logic resources and is faster than comparable methods. Although the functionality of the proposed solution has been demonstrated for the Altera Stratix family, the same principle and design methodology can be used for all recent high-performance ASICs or FPLDs that include an on-chip reconfigurable analog PLL.

The generator is developed for embedded cryptographic applications which help to increase the system security but it can also be used in a wide range of other applications.

References

- [1] Fischer V., et al. High performance true random number generator in Altera Stratix FPLDs. "Field-Programmable Logic and Applications," 14th International Conference, FPL 2004, Antwerp, Belgium, August 30-September 1, 2004, LNCS 3203, Springer, Berlin, Germany.
- [2] Fischer V., et al. Simple PLL-based based true random number generator for embedded digital systems. 7th IEEE workshop Design and Diagnostic of Electronic Circuits and Systems, April 18-21, 2004, Stara Lesna, Slovakia.
- [3] Drutarovsky M. et al. Implementation of hardware (true) random number generators (TRNGs) in reconfigurable devices. CryptArchi 2003, January 9-11, 2003, Saint-Etienne, France.
- [4] Kohlbrenner P. and Gaj K., An embedded true random number generator for FPGAs. FPGA'04, February 22-24, 2004, Monterey, California, USA.
- [5] Altera Technical Brief 70, Jitter Comparison Analysis: APEX 20KE PLL vs. Virtex-E DLL ver. 1.1, January 2001.
- [6] Altera Application Note 115, Using the clocklock & clockboost PLL features in APEX devices ver 2.6, November 2003.
- [7] Altera Application Note 282, Implementing PLL reconfiguration in Stratix & Stratix GX devices ver 2.0, December 2005.
- [8] Altera Data Sheet: NIOS Development Board Reference Manual, Stratix Professional Edition ver 1.1, December 2003.
- [9] Altera Application Note: Design debugging using signaltap II Embedded logic analyzer, June 2004.
- [10] Rukhin A. et al., A statistical test suite for random and pseudorandom number generators for cryptographic applications, NIST Special Publication 800-22, May 15, 2001.

Accelerating Graph Algorithms with Priority Queue Processor

Ch'ng Heng Sun¹, Chew Eik Wee², Nasir Shaikh-Husin³, Mohamed Khalil Hani⁴

VLSI-ECAD Research Laboratory (P04-Level 1),
Microelectronic and Computer Engineering Department (MiCE),
Faculty of Electrical Engineering (FKE),
Universiti Teknologi Malaysia.

¹chnghengsun@yahoo.com.sg

²eikweechew@hotmail.com

³nasirsh@utm.my

⁴khalil@fke.utm.my

Abstract

Graphs are a pervasive data structure in computer science, and algorithms working with them are fundamental to the field. Of the various graph algorithms, techniques for searching a graph are the heart of graph algorithms. Many graph algorithms are organized as simple elaborations of basic graph searching algorithms. For the searching of a graph, Priority Queue is used to maintain the tentative search result and choice of priority queue implementation would significantly affect the run-time and memory consumption of a graph algorithm. In this work, we demonstrate how to accelerate graph algorithms using priority queue processor. Dijkstra's algorithm is chosen as the target implementation, as many state-of-the-art graph algorithms use Dijkstra's algorithm at the heart of their computational engine. Assuming embedded hardware-software co-design, results show that our priority queue processor performs better than software implementation, and the run-time complexity of Dijkstra's algorithm is reduced from $O(n \lg n)$ in software implementation to $O(n)$ with our priority queue processor.

Keywords:

Graph Algorithms, Shortest Path Routing, Priority Queue, Priority Queue Processor

1. Introduction

Graphs are a pervasive data structure in computer science, and algorithms working with them are fundamental to the field, among others; Depth-First Search, Breadth-First Search, Topological Search, Spanning Tree Algorithms, Dijkstra's Algorithms, Bellman-Ford Algorithms, Floyd-Warshall Algorithm, etc. In real world applications, there exist many algorithms which are actually extended from these basic graph algorithms. For instance, in the field of VLSI physical design automation, there are many interesting computational problems defined in terms of graph; among others, Lee's Algorithm, Maze Routing Algorithms, Matching Algorithms, Min-Cut and Max-Cut Algorithms, Minimum Steiner Tree Algorithms, Span Minimum Tree Algorithms, Clock Skew Scheduling, Clock Net Synthesis, Critical Net Routing, etc.

A graph, $G = (V, E)$ consists of $|V|$ numbers of vertices/nodes and $|E|$ numbers of edges. Real world problems modeled in mathematical set can be represented as graphs, where elements in the set are represented by vertices, and the relation between any two elements are represented by edges. The run-time complexity and

memory-consumption of graph algorithms are expressed in terms of $|V|$ and $|E|$. A graph searching algorithm can discover much about the structure of a graph and many graph algorithms are organized as simple elaborations of basic graph searching algorithms [1]. Searching a graph means systematically following the edges of the graph so as to visit the vertices of graph. **For the searching of a graph, Priority Queue is used to maintain the tentative search result and choice of priority queue implementation would significantly affect the run-time and memory consumption of a graph algorithm** [2].

Priority Queue is an abstract data structure to maintain a set of elements, where all elements are arranged in accordance to their associate-priority. The associate-priority can be given as time-of-occurrence, level-of-importance, physical-parameters, delay/latency, etc, depending on target application. Two basic operations are supported by priority queue, namely (i) **INSERT(Q, x)**, which is generally referred to as **ENQUEUE** operation, and (ii) **EXTRACT(Q)** which is sometimes referred to as **DEQUEUE** operation. The performance of priority queue operations are measured in terms of n , where n refers to the total number of elements in the queue.

In many advanced algorithms where items/tasks are processed according to a particular order, priority queue has proven to be very useful. For *task-scheduling* on a multi-thread, shared-memory computer; priority queue is used to schedule and keep track of the prioritized pending processor tasks/threads. In the case of *discrete-event-simulation* (DES), priority queue is used where items in the queue are pending-event-sets, each with associated time-of-occurrence that serves as priority; many simulators, emulators, and synthesizers are designed based on DES concept. In the case of shortest-path based graph problem, priority queue has been used extensively in *QoS internet packet routing, weighted shortest path problem, multi-constrained routing, VLSI routing, PCB routing*, etc.

In this paper, choices of priority queue implementation will be discussed in section 2. Section 3 introduces the priority queue processor in this work to accelerate the graph algorithm. Section 4 illustrates the Dijkstra’s Algorithm. The embedded system architecture featuring the priority queue processor is described in section 5. Section 6 presents the result of this work.

2. Choices of Priority Queue

As shown in Figure 1, operation **INSERT(Q, x)** inserts a new element, *x* into priority queue, *Q*; meanwhile **EXTRACT(Q)** returns the element with the highest priority. Preferably, in some circumstances where the highest priority is given by a minimum value, the term **EXTRACT-MIN(Q)** is used instead of the more general term **EXTRACT(Q)**; whereas in other cases where the highest priority is the maximum value, the term **EXTRACT-MAX(Q)** is used. Hereinafter, **EXTRACT(Q)** is used interchangeably with **EXTRACT-MIN(Q)** or **EXTRACT-MAX(Q)**.

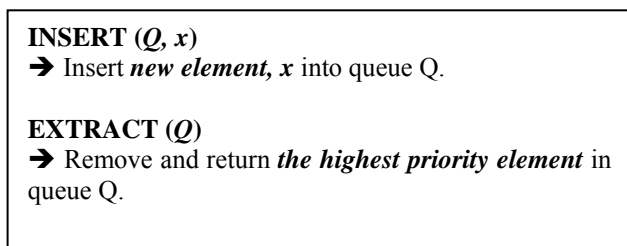


Figure 1. Basic Priority Queue Operations

For works done to implement a priority queue, either one of **EXTRACT-MIN(Q)** or **EXTRACT-MAX(Q)** function will be implemented, depending on what application is targeted. While software implementation could be easily modified to switch from **EXTRACT-MIN(Q)** to **EXTRACT-MAX(Q)**, or the other way round, it is not the case for full-custom hardware implementation. Anyway, this is not a big issue since maximum can be treated as reciprocal to the minimum, or vice-versa (maximum = 1/minimum). For instance, given a priority queue which only provides **INSERT(Q)** and **EXTRACT-MIN(Q)**, but the target-application needs **EXTRACT-**

MAX(Q), then simply solve it by inverting all the associate-priority (1/priority) before insertion into queue.

The simplest way to implement a priority is to keep an associate array mapping each priority to a list of items with that priority, see Figure 2. The priorities are held in static array which stores the pointers to the list of items assigned with that priority. Such implementation is static, for example, if the allowed priority ranged from 1 to 4,294,967,295 (32-bit) then an array of (4 Giga-length) * (size of pointer storage, i.e. 32-bit) is consumed, a total of 16 Gigabytes is gone, just to construct a priority data structure.

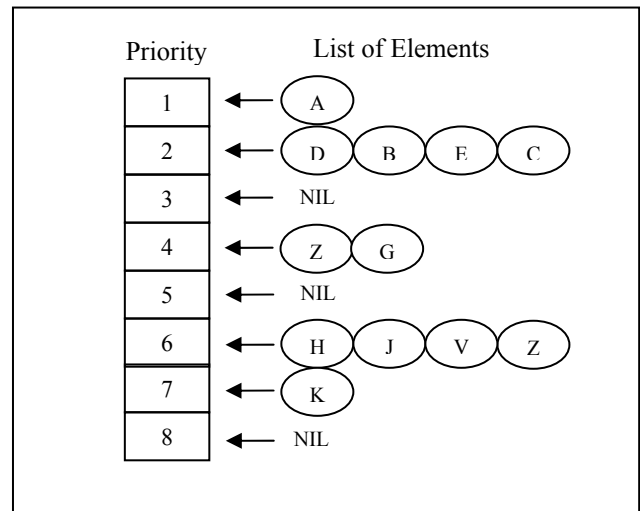


Figure 2. Simplest way to implement Priority Queue

A more flexible and practical way to implement a priority queue is to use dynamic array/queue, see Figure 3(a), which means the length of array does not depend on the range of priority. Each **INSERT(Q, x)** will extend the existing queue-length by one unit ($n \leftarrow n + 1$), put in the item with priority *x*, then sort the whole queue to ensure the highest priority item is ready if **EXTRACT(Q)** is invoked. Such sorting during insertion will take up to $O(n)$ computation run time. For extraction, since the whole queue already sorted during insertion, extraction takes constant $O(1)$ time.

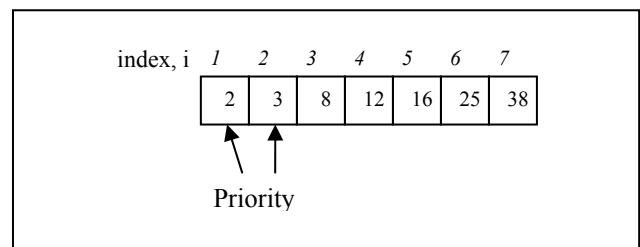


Figure 3(a). Priority Queue, implemented as array.

If a self-balanced Binary-Heap is used, see Figure 3(b), each **INSERT(Q, x)** or **EXTRACT(Q)** take $O(\lg n)$ run-time. In order to achieve better priority queue performance, many novel approaches have been taken to implement different structure of heap, i.e. Binomial-Heap, Fibonacci-Heap, Relaxed-Heap, Parallel-Heap, to name a few. A

developer has a number of choices with different access speed, memory consumption, and required hardware platform. Theoretically, Fibonacci-Heap Priority Queue (“FHPQ”) has better performance compared to other types of heap, especially for applications where the number of EXTRACT and DELETE operations are relatively small compared to INSERT and DECREASE-KEY operations; however, the drawback of Fibonacci-Heap is its complexity in practice with larger constant-time-factor.

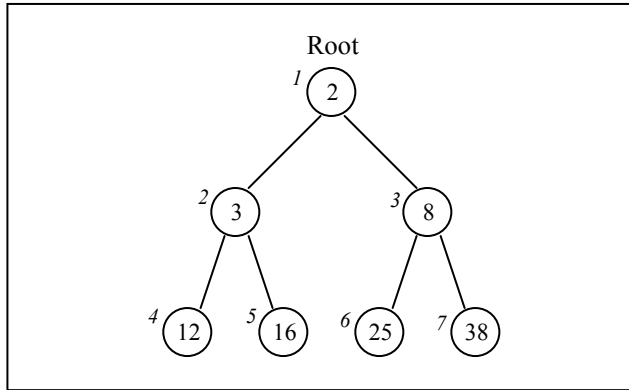


Figure 3(b). Priority Queue, implemented as heap. (in this case, Binary-Heap)

Due to the tremendous application of priority queue, there are a lot of researches done on how to accelerate the priority queue operation using custom hardware resources. There are two main categories of this work; (a) proposed full-custom parallel hardware design to accelerate array-based priority queue, (b) no specific hardware designed, but assuming PRAM (Parallel Random Access Machine) model, and propose new heap data structure executing on these PRAM. The former often achieves very high throughput and clocking frequency, favorite choice to high speed applications such as QoS network routing and real-time applications; whereas the latter often ignore the severe memory communication overhead and gives theoretical improvement in terms of run-time complexity.

Due to limited space, works in (b) is less related to our design work, it will be excluded, reader could refer to [3] for details. On the other hand, works in (a) includes *Binary Trees of Comparator* by [4]; such hardware priority queue is not efficient if the queue size is large, large fan-in for Binary-Tree also caused severe bus loading effect when it comes to physical implementation. Another approach to implement priority queue is using *FIFO Priority Queue* as proposed by [5] and [6], but the main drawback is the priority range is not flexible, just like the illustration in figure 2, thus not practical if the allowed priority range is large. *Shift Register* implementation of priority queue [6], [7] has better performance compared to Binary Trees of Comparators and FIFO Priority because the priority level and the queue size can be easily scaled, but Shift Register also suffers severe bus loading effect when it comes to real hardware implementation. To overcome the bus-loading effect of Shift Register, A hybrid design of *Systolic-Shift-Register* priority queue processor had been proposed by [8], which claims to reduce bus loading effect and could have

achieved better clocking rate. Having said that, all these designs are implemented on hardware high-speed network routers, they are different from embedded platform requirement; such as in our context, graph algorithm computation (or any other DES). For example, all these designs are targeting small number of priorities (i.e. 8-bit), but in the case for graph algorithm; the data which represents the priority can be in any range (i.e. 32-bit); furthermore, for embedded system implementation, those device are practically not accessible to us.

3. Systolic Array Priority Queue Processor

In a programming model, priority queue can be viewed as an array; when new element is inserted, it will be compared with all other elements in queue, in order to determine where to slip in this new element, in such a way that priority-order of the queue is always maintained. This approach is known as insertion-sort priority queue. Derived from the insertion-sort algorithm, $O(n)$ run-time per operation is needed for a priority queue with n elements.

Given an all-pair-single-source shortest-path graph problem to be solved with the Dijkstra’s algorithm, the run-time complexity is $O(n^2)$ if insertion-sort priority queue (hereinafter called “ISPQ”) is used; the run-time can be improved to $O(n \lg n)$ if the priority queue is implemented as binary-heap priority queue (“BHPQ”). This shows that different choices of priority queue implementation could significantly affect the overall performance of graph algorithm, especially when the queue size is very large.

An in-house-developed hardware, Systolic Array Priority Queue Processor (“SAPQ”) could further reduce the run-time complexity of the said graph algorithm to $O(n)$. The SAPQ, benefited from parallelizing a modified version of insertion sort algorithm supports both **INSERT**(Q, x) and **EXTRACT**(Q) in constant $O(1)$ run time. Improved from the Systolic-Shift-Register by [8], the architecture of SAPQ is much simpler, localizing control-unit and datapath for each processing element (PE) making it highly pipelined, easy for cascading, and with no bus-loading effect.

The SAPQ consists of an array of *identical* processing elements (PEs), with each PE holding a single queue-element. The identical-PE feature of SAPQ makes it very flexible, where a *parameterized* design always allows generation of any queue-size in a FPGA environment. A very large queue-size can be achieved by cascading multiple FPGAs on a single board or distributed through multiple boards. The SAPQ employs n number of PEs for worst-case n -size priority queue. Each PE is interconnected to only its immediate-neighbours. Figure 4 illustrates the top-level architecture of SAPQ. Of all the PEs, only the left-most PE will communicate with the outside world, meaning the new element is inserted to PE1, as well as the highest priority element will be extracted from PE1.

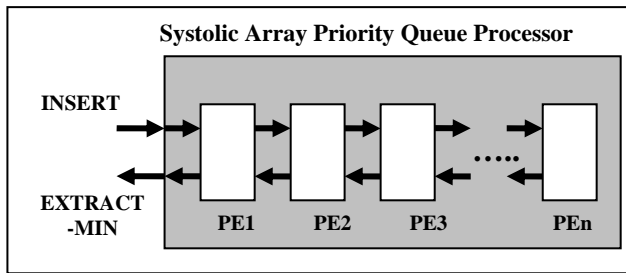


Figure 4. Systolic Array Priority Queue Processor

Each INSERT and EXTRACT-MIN operation on SAPQ completes in 3 clock cycles. The design of SAPQ allows, for each queue element, the element-ID in 32-bit, so as the associate-priority in 32-bit. The element-ID can be treated as associate-pointer which points to a block of satellite-data of that element.

The SAPQ design is compiled and targeted to ALTERA Stratix and Cyclone II FPGA devices. The synthesis shows that the design can obtain 240 MHz clocking frequency in Stratix and 175 MHz in Cyclone II. The design achieved 5.12 Gbps throughput rate for Stratix and 3.73 Gbps in Cyclone II. For details of SAPQ design, refer to [9].

4. Dijkstra's Graph Algorithm

All shortest path based graph algorithms follow a sequence of predefined graph-search procedures to ensure the shortest paths are found (if any) within the minimum period of time. In this paper, Dijkstra's algorithm is selected for illustration purpose. Refer the pseudo-code in Listing 1.

Given a graph $G = [V, E]$, where $V[G]$ denotes the set of vertices and $E[G]$ denotes the set of edge-weights. We use s to denote the source-vertex, for any two adjacently-connected vertices, $v = Adj[u]$ or $u = Adj[v]$. $d[u]$ denotes total-path-length from s to u , whereas $d[v]$ denotes total-path-length from s to v . Given $w(u, v)$ denoted the edge-weights from u to v , then $d[v] = d[u] + w(u, v)$. Specifically for shortest path algorithm, S is the set used to hold the nearest vertex from source, $\pi[v]$ is used to hold the precedent-vertex of v . Upon complete execution of algorithm, the shortest path from s to v can be traced by dereference $\pi[v]$ backward to source, and the total-path-length is given by $d[v]$.

Dijkstra's algorithm begins with source-node, where all nodes adjacent to source will be scanned and the edge-weights inserted into priority queue. Then the edge with minimum-weight will be extracted from priority queue, and the node in which this minimum-weight-edge heading-to becomes active. The same scanning mechanism continues at this current active node, with insertion to priority queue if there is no identical entry in queue, or relaxation (decrease-key) if identical entry already exists in queue. The procedure continues until all nodes have been visited.

```

DIJKSTRA( $G, w, s$ ) {
  for (each vertex  $v \in V[G]$  and  $v \neq s$ ) {
     $d[v] \leftarrow \infty$ 
     $\pi[v] \leftarrow \text{NIL}$ 
  }
   $d[s] \leftarrow 0$ 
  INSERT( $Q, s, d[s]$ )
   $S \leftarrow \emptyset$ 
  do {
    ( $u, d[u]$ )  $\leftarrow$  EXTRACT-MIN( $Q$ )
     $S \leftarrow S \cup \{u\}$ 
    for (each vertex  $v \in Adj[u]$ ) {
      if ( $d[v] = \infty$ ) {
         $d[v] \leftarrow d[u] + w(u, v)$ 
         $\pi[v] \leftarrow u$ 
        INSERT( $Q, v, d[v]$ )
      }
      elseif ( $d[v] > d[u] + w(u, v)$ ) {
         $d[v] \leftarrow d[u] + w(u, v)$ 
         $\pi[v] \leftarrow u$ 
        DECREASE-KEY( $Q, v, d[v]$ )
      }
    }
  } (while  $Q \neq \emptyset$ )
}

```

Listing 1. Pseudo-code of Dijkstra's Algorithm

```

DECREASE-KEY ( $Q, x, \text{new}(x\_priority)$ )
   $\rightarrow$  Find element  $x$  in queue,  $Q$ ;
  If ( $\text{new}(x\_priority)$  dominates  $\text{old}(x\_priority)$ )
     $x\_priority \leftarrow \text{new}(x\_priority)$ .

```

Figure 5. Decrease-Key function

Notice that besides the basic priority queue operations in Figure 1, the graph algorithm needs one additional function, the DECREASE-KEY. The DECREASE-KEY function is very important because it is necessarily for any graph algorithm to perform **RELAXATION** (see [1]). Recall each element x in queue Q comes with its ID (here, denoted as x_ID) and its associate-priority (here, denoted as $x_priority$). Refer to Figure 5, when it comes to a condition where the associate-priority, $x_priority$ of an element x needs update, the DECREASE-KEY is invoked to find that element x in queue; if the new-associate-priority-of- x , $\text{new}(x_priority)$ possesses higher priority than the old-associate-priority-of- x , $\text{old}(x_priority)$ (the one exist in queue), then replace that associate-priority-of- x . Such property is called **dominancy**, and it is the key to understand graph algorithms.

```

DIJKSTRA_MODIFIED( $G, w, s$ ) {
  for (each vertex  $v \in V[G]$ ) {
     $d[v] \leftarrow \infty$ 
     $\pi[v] \leftarrow \text{NIL}$ 
  }
   $d[s] \leftarrow 0$ 
  INSERT( $Q, s, d[s]$ )
   $S \leftarrow \emptyset$ 
  do {
    do {
      ( $u, temp$ )  $\leftarrow$  EXTRACT-MIN( $Q$ )
    } (while  $d[u] \neq temp$ )
     $S \leftarrow S \cup \{u\}$ 
    for (each vertex  $v \in Adj[u]$ ) {
      if ( $d[v] = \infty$ ) {
         $d[v] \leftarrow d[u] + w(u, v)$ 
         $\pi[v] \leftarrow u$ 
        INSERT( $Q, v, d[v]$ )
      }
      elseif ( $d[v] > d[u] + w(u, v)$ ) {
         $d[v] \leftarrow d[u] + w(u, v)$ 
         $\pi[v] \leftarrow u$ 
        INSERT( $Q, v, d[v]$ )
      }
    }
  } (while  $Q$  not empty)
}

```

Listing 2. Modified version of Dijkstra's Algorithm

All software implementation of priority queue in heap (Binary-Heap, Binomial-Heap, and Fibonacci-Heap) could easily incorporate all three functions needed in finding the shortest path: INSERT, EXTRACT-MIN, and DECREASE-KEY. However, neither hardware implementation of priority queue ever provides DECREASE-KEY function; this function is possible to be implemented but often ignored due to additional hardware resources required. The design of SAPQ also does not incorporate DECREASE-KEY function. Hence, modification on the targeted Dijkstra's graph algorithm is needed, so that the algorithm does not invoke DECREASE-KEY directly for relaxation, but indirectly perform relaxation using the only available **INSERT**(Q, x) and **EXTRACT-MIN**(Q).

The modified version of Dijkstra's algorithm is presented in Listing 2, which only invoked INSERT and EXTRACT-MIN functions. The sequence of execution remains except (i) when supposed DECREASE-KEY is needed, INSERT is invoked instead; this causes the queue size to actually grow by one, and there is one entry in queue

which is no longer valid. (ii) During EXTRACT-MIN, the returned queue-entry will be checked for validity, until a valid entry is returned. Such modification retains the relaxation property of graph algorithm, except the priority queue size at some instance might grow larger than algorithm in Listing 1. Anyway, such exception does not have profound effect on SAPQ since the operation run-time is constant, unlike ISPQ, BHPQ, or other heaps which the operation run-time depends on queue-size.

5. Embedded System Implementation

Targeted for embedded system implementation, SAPQ will serve as co-processor, to off-load the recursive priority queue access and maintenance from the host processor. Our implementation, assuming a general purpose processor, NIOS, to serve as host processor, executing (modified) Dijkstra's algorithm, with the priority queue access and maintenance fully handled by the SAPQ. The whole system is prototyped on a FPGA as illustrated in Figure 6. The Avalon Interface Unit is designed to handle the communication protocol between host processor and SAPQ. The embedded system could be deployed to handle all kinds of application which utilize priority queue. The Dijkstra's algorithm executed by the host processor can be replaced by other algorithms such as discrete-event simulation, global-positioning-system, mobile navigating, etc.

Due to the limitation of Avalon bus bandwidth, which is 32-bit width, the SAPQ does not actually execute at its 64-bit interfacing capability. An optimal SAPQ could be achieved by having it implemented in higher-end communication bridge, such as PCI-64 as the system bus.

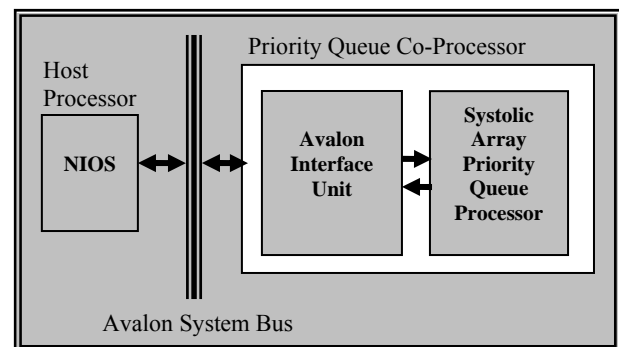


Figure 6. Embedded System on FPGA

6. Result

SAPQ with queue-size of 200 entries is implemented on the said embedded system; this queue-size is constrained by the logic resources of this particular FPGA. Several software priority queues (ISPQ, BHPQ, and FHPQ) are used to compare the performance of SAPQ. Having the SAPQ actually running at lower than optimal speed (50 MHz compared to the maximum allowed is 240 MHz), narrower bus bandwidth (32-bit compared to allowed 64-bit), and high redundancy in cycles per operation incurred by the host processor (44 cycles for INSERT and 52 cycles for EXTRACT; compared to actually 3 cycles designed for

INSERT and 3 cycles designed for EXTRACT), the gain achieved in terms of number of cycles spent per operation is still significant.

(Queue Size = n)	Resource Utility		Run Time Complexity, n	
	# of Processor	# of Memory Storage	INSERT	EXT-MIN
Systolic Array Priority Queue, SAPQ	O(n)	NIL	O(1)	O(1)
Binary-Heap Priority Queue, BHPQ	O(1)	O(n)	O(lg n)	O(lg n)
Insertion-Sort Priority Queue, ISPQ	O(1)	O(n)	O(n)	O(1)
Fibonacci-Heap Priority Queue, FHPQ	O(1)	O(n)	O(1)	O(lg n)*

Table 1: Comparisons in terms of run-time complexity.

(Queue size limited to n=200)	WORST CASE (number of cycles)		SPEED UP GAIN (Achieved by SAPQ)	
	INSERT	EXT-MIN	INSERT	EXT-MIN
Systolic Array Priority Queue, SAPQ	44	52	1.00	1.00
Binary-Heap Priority Queue, BHPQ	753	1185	17.11	22.79
Insertion-Sort Priority Queue, ISPQ	13674	65	310.77	1.25
Fibonacci-Heap Priority Queue, FHPQ	799	44262	18.16	851.19

Table 2: Comparison in terms of processor cycles.

Refer to Table 1, the performance of priority queue depends on the queue-size, n . In Table 2, the worst case INSERT is for inserting entries so that the queue is full. The worst case EXT-MIN is for extracting the minimum entry when the queue is full.

Compared to BHPQ, the speed up gain achieved by SAPQ is more than expected. Theoretically, one could expect $\lg n$ speed up (which in this case $n = 200$, $\log_2 200 = 7.6$) by SAPQ; this real implementation achieved 17 times and 22 times worst-case speed up for INSERT and EXTRACT operation. This is because software implementation suffers severe memory communication overhead, where tremendous cycles are spent to access heap data structure stored in memory; whereas SAPQ, having all the elements stored in registers, such communication drawback is eliminated.

* For FHPQ, the $O(\lg n)$ for EXT-MIN is worst-case amortized-time, refer [1] for amortized analysis.

Next, compared to ISPQ, SAPQ obtained 310 speed gains for worst-case INSERT and 1.25 gains for EXTRACT. Notice the 310 times speed gain is also greater than theoretical expectation; it is also due to the advantage of hardware implementation which eliminates the memory communication overhead.

Lastly, comparison is made between the theoretically most efficient priority queue, the FHPQ with our SAPQ; the result is very impressive. Both FHPQ and SAPQ claims $O(1)$ run-time complexity for INSERT, this implementation shows 18 times gain achieved using SAPQ because constantly large cycles are spent by FHPQ to handle a bunch of pointer manipulation. Similarly, for EXTRACT operation, the speed-up gain achieved by SAPQ is over 851 times.

The above reported worst case speed-up gain could be more if larger SAPQ were to be implemented (i.e. queue-size of 1000). The main drawback of Systolic Array Priority Queue Processor is the logic resources consumed, when other software implementation of software priority queues only take space in random access memory. Anyway, for real-time applications where the speed is top-priority, the drawback in logic consumption is a worthy trade-off.

References

- [1] T. H. Cormen, C. E. Leiserson, R. L. Rivest, Clifford Stein, Introduction To Algorithms, 2nd Edition, The MIT Press, McGraw-Hill Book Company, 2001.
- [2] S. S. Skiena, The Algorithm Design Manual, Springer-Verlag, New York, 1997
- [3] Ch'ng Heng Sun, Mohamed Khalil Hani, "Design of a Graph Processor for VLSI Routing", Master Research Proposal, FEB. 2006.
- [4] D. Picker and R. Fellman, "A VLSI Priority Packet Queue with Inheritance and Overwrite," IEEE Trans. Very Large Scale Integration Systems, vol. 3, no. 2, pp. 245-252, June 1995.
- [5] R. Brown, "Calendar Queues: A Fast $O(1)$ Priority Queue Implementation for the Simulation Event Set Problem," Comm. ACM, vol. 31, no. 10, pp. 1220-1227, Oct 1988.
- [6] J. Chao, "A Novel Architecture for Queue Management in the ATM Network," IEEE J. Selected Areas in Comm., vol. 9, no. 7, pp. 1110-1118, Sept. 1991.
- [7] K. Toda, K. Nishida, E. Takahashi, N. Michell, and Y. Tamaguchi, "Design and Implementation of a Priority Forwarding Router Chip for Real-Time Interconnect Networks," Int'l J. Mini and Microcomputers, vol. 17, no. 1, pp. 42-51, 1995.
- [8] S.W. Moon, J. Rexford, K.G. Shin, "Scalable Hardware Priority Queue Architectures for High-Speed Packet Swiches", IEEE Trans. On Computers, vol. 49, no. 11, Nov. 2000.
- [9] Ch'ng Heng Sun, Mohamed Khalil Hani, "Systolic Array Priority Queue Processor", Research Report, APR. 2006.

Crypto Embedded System for Electronic Document

Iliasaak Ahmad¹, Norashikin M. Thamrin¹, Mohamed Khalil Hani¹

¹VLSI-ECAD Research Laboratory, Microelectronic and Computer Engineering Department (MiCE)
Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
Tel: +60-7-5535223, Fax: +60-7-5566272, E-mail: ilyasak_ahmad@yahoo.com,
norashikin.mthamrin@gmail.com, khalil@fke.utm.my

Abstract

In this paper, a development of low-cost RSA-based Crypto Embedded System targeted for electronic document security is presented. The RSA algorithm is implemented in a re-configurable hardware, in this case Field Programmable Gate Array (FPGA). The 32-bit soft cores of Altera's Nios RISC processor is used as the basic building blocks of the proposed complete embedded solutions. Altera's SOPC Builder is used to facilitate the development of crypto embedded system, particularly in hardware/software integration stage. The use of Cryptographic Application Programming Interface (CAPI) to bridge the application and the hardware, and the associated communication layer in the embedded system is also discussed. The result obtained shows that the crypto embedded system provides a suitable compromise between the constraints of speed, space and required security level based on the specific demands of targeted applications.

Keywords:

Embedded System, Public Key Cryptography, FPGA, HW/SW Co-design

1. Introduction

Cryptography has gained an important role in today's information security problems. Security of the system can be enhanced if it is embedded in a re-configurable hardware. Such implementation is harder to tap, decompose, and attack, in general.

The protocols in public key cryptography like RSA, El-Gamal, etc, are excellent examples for implementing HW/SW co-design concept. Public key cryptography is based on the difficulty of factoring large numbers. To increase the operation speed, the algorithm is most often realized as a hardware component based on a parallel array of processing elements [1][2]. The hardware structures are generally fast enough, but not suitable for algorithm sequencing and they cannot be adapted to algorithm changes. Software, on the other hand, adapts itself easily but it is much slower and less secure. Thus implementing cryptographic algorithms in re-configurable hardware/SOC offers the best solution: it can consist of an embedded processor, one or more coprocessors, and software.

In this paper 1024-bit RSA algorithm is implemented. Due to hardware resource constraint, the encryption and verification modules are implemented as embedded code, while decryption and signing operations are performed in hardware. Chinese Remainder Theorem (CRT) is deployed not only to speed up decryption and signing operations, but also to utilize the 512-bit RSA co-processor designed in [3].

This paper is organized as follows: Section 2 covers the fundamental concept of RSA algorithm. The design of crypto embedded system is discussed in Section 3. Section 4 looks at the verification of the crypto embedded system and its performance. We discuss, in brief, the use of Cryptographic Application Programming Interface (CAPI) as high-level

interface to the crypto embedded system in Section 5. Finally, concluding remarks is presented in the final section.

2. Overview of RSA Algorithm

As reported in [4], the most widely used public-key algorithm is RSA algorithm. This is due to the fact that, RSA can provide both confidentiality and digital signatures using the key-pair and under the same mathematical operation. Figure 1(a), 1(b), and 1(c) summarize RSA algorithm.

1. Generate 2 primes, p and q randomly, where $p \neq q$
2. Calculate M , where $M = p * q$
3. Calculate $\phi(M)$, where $\phi(M) = (p - 1)(q - 1)$
4. Generate E (public exponent) that fulfills $1 < E < \phi(M)$ and $\text{GCD}(\phi(M), E) = 1$
5. Calculate D (private exponent), where $D = E^{-1} \text{Mod } \phi(M)$

Figure 1(a). RSA Key-Pair Generation

$$\begin{aligned} \text{Plaintext (P)} &< M \\ \text{Ciphertext (C)} &= P^E \text{ Mod } M \end{aligned}$$

Figure 1(b). RSA Encryption/Verification

$$\begin{aligned} \text{Ciphertext (C)} \\ \text{Plaintext (P)} &= C^E \text{ Mod } M \end{aligned}$$

Figure 1(c). RSA Decryption/Signing

3. Design of Crypto Embedded System

An important aspect in embedded system design is partitioning the overall system into hardware and software components. This involves the physical partitioning of functionality into hardware or software, and it is influenced by system requirement, availability of device resources, IP core, and execution time. Figure 2 depicts the generic embedded HW/SW design flow.

The architecture of the crypto embedded system is shown in Figure 3. It has a processor core, on-chip memory, a co-processor, UART communication, and internal system bus. Nios [5], a 32-bit soft core from Altera, is used as the processor, while Avalon Bus [6] is used to enable communication between processor and RSA co-processor. The RSA co-processor is designed to perform the intensive part of computation of RSA algorithm. Nios processor is used to implement the more control intensive, parameterizable portions of RSA algorithm like embedded encryption module, and some parts of decryption module. SHA-1 has also been implemented in Nios (to be used with digital signature operation).

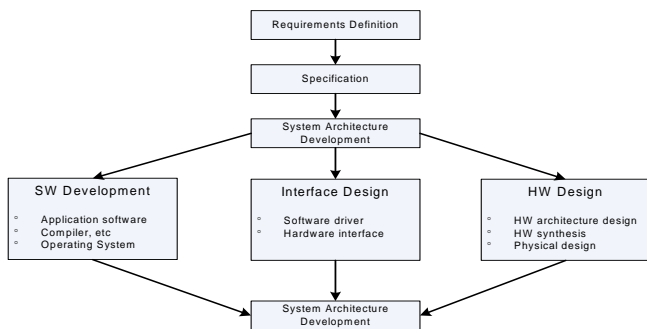


Figure 2. Generic Embedded HW/SW Design Flow

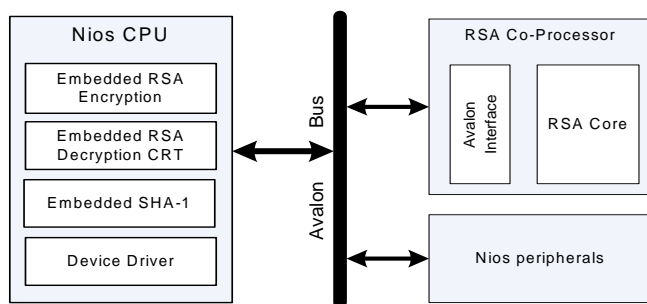


Figure 3. System Architecture of Crypto Embedded System

3.1 Interfacing RSA Co-processor with Embedded Processor

In this work, 512-bit RSA co-processor designed in [3] to perform 1024-bit RSA operation is used. This co-processor is designed to handle intensive modular arithmetic computations. It is, however, beyond the scope of this paper to discuss the implementation of this co-processor in detail.

To interface this co-processor with Nios processor, Avalon Bus is used as the bus system. Avalon Bus is a

simple bus architecture designed for connecting on-chip processor and peripherals together into a system on a programmable chip. It is an interface that specifies the port connections between master and slave components, and also specifies the timing by which the components communicate [6]. Avalon Bus transactions transfer a single byte, half-word, or word (8, 16, or 32-bits) between a master and slave peripheral.

The RSA co-processor used in this work uses Avalon slave transfer mode that accept Avalon bus transfer from master port, which is Nios processor.

3.2 HW/SW Integration

Tools are available to integrate the processor, co-processor, and peripherals to become a complete embedded system. In this work, we used Altera SOPC Builder [7]. See Figure 4 for an illustration of SOPC Builder HW/SW integration flow. It is the interest of this paper to explore in greater details the Software Development portion of the flow.

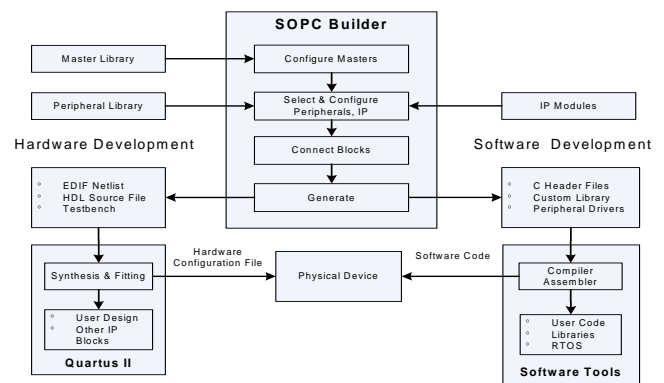


Figure 4. SOPC Builder HW/SW integration flow [8]

To generate the software development kit (SDK) for the embedded system, all the peripherals and IP core used must be added and configured, see Figure 5. The RSA Co-processor is added here as user-defined interface. The base address of this co-processor is set to 0x0900900. Since Altera Apex EP20K20EFC484-2X development board is used, the clock frequency is set to 33.33 MHz. Other configuration parameters include assignment of interrupt priorities, and the setup, hold and wait requirements of each peripheral.

On the completion of generation process, SOPC Builder generates the hardware and software driver file called excalibur.h. Excalibur.h includes all the software interfaces for all the blocks in the embedded system. Apart from that, excalibur.h also includes the address for all registers and memories inside the SOPC Builder as well as associated software application programming interfaces (APIs) for IP blocks that include APIs. Figure 6(a) and 6(b) show the excerpts taken from excalibur.h.

3.3 Embedded Software Development

Based on the previous section, the SOPC builder generates a custom SDK. The SDK contains the memory map and data structures for accessing hardware components in the system.

It also provides routines for accessing the peripherals like UART. So this SDK can be used to communicate easily with fundamental system components and custom IP cores.

Application programming interfaces are developed using C language to control the RSA Co-processor operations. The interfaces are:

- RSA_OperandM()
- RSA_OperandE()
- RSA_OperandR()
- RSA_OperandX()
- RSA_Ouput()
- RSA_MonMult_vec()
- Compute_R()

Besides that, embedded code for encryption and decryption with CRT are also coded. Figure 7 illustrates the APIs and the embedded software.

A device driver for the crypto embedded system is also developed to enable communication between crypto embedded system and external world like personal computer. The flow of the device driver can be described as state diagram. See Figure 8.

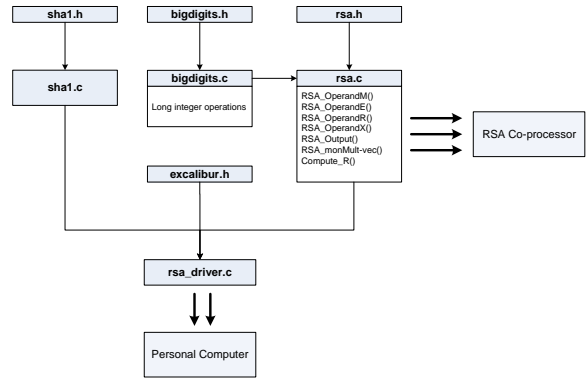


Figure 7. APIs and Embedded Code

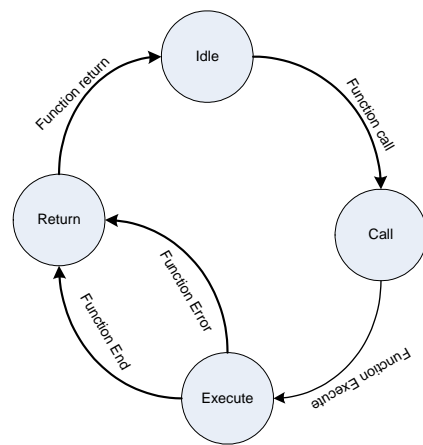


Figure 8. Device driver state diagram

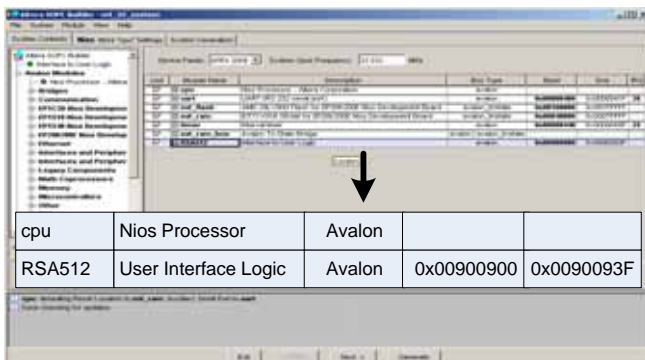


Figure 5. Configuring Crypto Embedded System Using SOPC Builder

```

// The Memory Map
#define na_cpu ((void *) 0x00000000) // altera_nios
#define na_cpu_base 0x00000000
#define na_uart ((np_uart *) 0x00000400) // altera_avalon_uart
#define na_uart_base 0x00000400
#define na_uart_irq 26
#define na_timer ((np_timer *) 0x00000440) // altera_avalon_timer
#define na_timer_base 0x00000440
#define na_timer_irq 25
#define na_ext_sram ((void *) 0x00040000) // altera_nios_dev_board_sram32
#define na_ext_sram_base 0x00040000
#define na_ext_sram_end ((void *) 0x00080000)
#define na_ext_sram_size 0x00040000
#define na_ext_flash ((void *) 0x00100000) // altera_nios_dev_board_flash
#define na_ext_flash_base 0x00100000
#define na_ext_flash_end ((void *) 0x00200000)
#define na_ext_flash_size 0x00100000
#define na_RSA512 ((np_usersocket *) 0x00900900) // altera_avalon_user_defined_interface
#define na_RSA512_base 0x00900900
    
```

Figure 6(a). Excerpt from excalibur.h: Memory Map

```

// Structures and Routines For Each Peripheral
// Nios CPU Routines
void nr_installmanager(void); // called automatically by nr_setup.s
void nr_delay(int milliseconds); // approximate timing based on clock speed
void nr_zerorange(char *rangeStart,int rangeByteCount);
void nr_jumptoreset(void);
// Default UART routines
void nr_txchar(int c);
void nr_txchar2(int c, int channel);
void nr_txstring(char *s);
int nr_rxchar(void);
    
```

Figure 6(b). Excerpt from excalibur.h: APIs

4. Result

To evaluate and verify software/hardware modules in this solution and its impact on the speed of the RSA operation, test application is developed. The test application is run on Nios processor.

The test vector is obtained from National Institute of Standards and Technology (NIST) [9] and hard-coded in the test program. For this test we have predefined the value of public key to be ‘17’, and this value can be easily changed according to the user’s requirement later. Table 1 shows the timing of the RSA operations (CRT is deployed in RSA decryption).

Table 1. Execution times of RSA on Altera APEX EP20K200EFC484-2X clocked at 33.33 MHz

Operation	Time (ms)
Encryption (software)	10
Decryption (software with CRT)	18338
Decryption (hardware with CRT)	111

The result obtained here is also compared with ARM SecurCore [11], one of the commercial products available in the market. Table 2 below summarizes the comparisons that are made. Based on that table, we can see that the result obtained is quite competitive in terms of performance.

Table 2. Comparison with commercial product

Specifications	Our design	ARM SecurCore
KeySize	1024	1024
Clock	33.33 Mhz	20 MHz
Decryption (with CRT)	111 ms	330 ms
Encryption (software)	10 ms	-

5. Cryptographic Application Programming Interface (CAPI)

Modern embedded systems are built using various techniques that provide flexibility and reliability. One of the most important techniques centres on the use of applications programming interface.

An application-programming interface (API) is basically a well-defined boundary between two system components that isolates a specific process or a set of services. For example, it is quite common now for an application to interact with e-mail service using e-mail API like MAPI (Microsoft), VIM (Lotus), and others. In such cases, the API defines a set of services that allow an application to retrieve or submit mail messages from or to the mail server.

A cryptographic application programming interface (CAPI), like other APIs, is an API specifically designed to support cryptographic functions. Technically, a CAPI would provide and interface to a set of cryptographic services such as encryption/decryption, digital signatures/verification, key generation, etc. Figure 9 depicts the relation between CAPI and crypto services.

A simple and easy to use CAPI called *myCAPI* is developed for the crypto embedded system presented in this paper. *myCAPI* has a set of well-defined APIs that enable application developers integrate crypto embedded system services into their application. The list of available APIs is listed in Table 3.

In addition to *myCAPI*, the crypto embedded system services can also be called-up through Microsoft CryptoAPI interface [10]. This integration would benefit the application developers. Since Microsoft CryptoAPI has been defined as one of the standard CAPI, application that utilizes Microsoft CAPI can access multiple cryptographic implementations through a single interface, see Figure 9.

6. Conclusion

In this paper, the design of crypto embedded system targeted for electronic document security has been presented. The crypto embedded system is implemented in re-configurable hardware, which is FPGA. Altera CAD tool, SOPC Builder

is used to facilitate and demonstrate hw/sw design and development flow. The result obtained shows that the crypto embedded system provides a suitable compromise between the constraint of speed, space and required security level based on the specific demands of targeted applications.

Table 3. *myCAPI* APIs

API	Description
utmGenKeyPair()	Key-pair generation
utmRSASigning()	RSA digital signatures
utmRSAVerification()	RSA digital signatures verification
utmRSAEncryption()	RSA encryption
utmRSADecryption()	RSA decryption

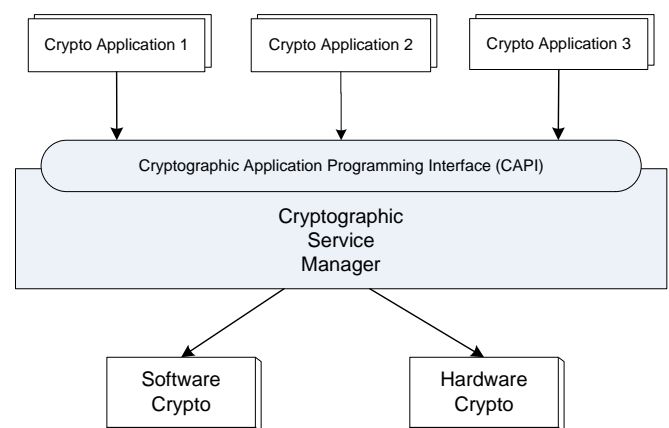


Figure 9. CAPI Architecture

References

- [1] A. F. Tenca and C. K. Koc. A scalable architecture for Montgomery multiplication. In C.K. Koc and C. Paar, editors, Cryptographic Hardware and Embedded Systems, number 1717 in Computer Science, pages 94-108, Berlin, Germany, 1999. Springer Verlag.
- [2] M. Drutarovský, V. Fischer, and M. Šimka. Two Implementation Methods of Scalable Montgomery Coprocessor Embedded in Reconfigurable Hardware. Cryptographic Hardware and Embedded Systems 2003.
- [3] Paniandi, A., 2005. A Hardware Implementation of RSA Co-processor for Resource Constrained Embedded Systems. Master Dissertation, Faculty of Electrical Engineering, Uni. Teknologi Malaysia.
- [4] B. Schneier. Applied Cryptography: Protocol, Algorithm and Source Code in C. 2nd Edition, John Wiley & Sons Inc, NY. 1996.

- [5] Nios Soft Core Embedded Processor, <http://www.altera.com/nios>
- [6] Avalon Bus Specification, <http://www.altera.com/literature/manual/>
- [7] Altera SOPC Builder, <http://www.altera.com/products/software/products/sopc/>
- [8] Ekas, P; Jentz B, Fall 2003. Developing and integrating FPGA coprocessors. Embedded Computing Design.
- [9] Keller. S. S, 2004. The RSA Validation System (RSAVS). National Institute of Standards and Technology (NIST). USA.
- [10] Microsoft Developer Network, <http://msdn.microsoft.com/library/>
- [11] ARM Products and Solutions – Core Type, <http://www.arm.com/products/CPUs/securcore.html>

Nano-scale VLSI Clock Routing Module based on Useful-Skew Tree Algorithm

Chew Eik Wee¹, Ch'ng Heng Sun², Nasir Sheikh-Husin³, Mohamed Khalil Hani⁴

VLSI-ECAD Research Laboratory (P04-Level 1),
Microelectronic and Computer Engineering Department (MiCE),
Faculty of Electrical Engineering (FKE),
Universiti Teknologi Malaysia.

¹Chew Eik Wee (eikweechew@hotmail.com)

²Ch'ng Heng Sun(chnghengsun@yahoo.com.sg)

³Nasir Shaikh-Husin(nasirsh@utm.my)

⁴Mohamed Khalil Hani (khalil@fke.utm.my)

Abstract

Clock routing is critical in nano-scale VLSI circuit design. Clock routing needs to be precise to minimize circuit delay. Clock signals are strongly affected by technology scaling, the long global interconnect lines become highly resistive as line dimensions are decreased. The control of clock skew can also severely limit the maximum performance of the entire system and create catastrophic race conditions in which an incorrect data signal may latch within a register. Thus, we propose a clock routing synthesis module that applies non-zero skew (or called useful-skew) method to reduce the system-wide minimum clock period to improve the performance of synchronous digital circuit. We implemented Useful-Skew Tree (UST) algorithm which is based on the deferred-merge embedding (DME) paradigm, as the clock layout synthesis engine. The synthesis module is integrated with the UTM in-house design graph accelerator to enhance the computation performance. The novel contribution of this work is clock skew scheduling is performed simultaneously with clock tree routing. This way, the computation result of proposed synthesis module can generate a clock signal distribution routing path with minimum wire length and, ensures the reliability of data synchronization for nano-scale VLSI design.

Keywords:

Clock Routing, Synthesis Module, Useful Skew, Clock Skew Scheduling, Useful-Skew Tree.

1. Introduction

In a synchronous digital system, the clock signal is used to define a time reference for the movement of data within the system. Since this function is vital to the operation of a synchronous system, much attention has been given to the characteristics of these clock signals and the routing path used in their distribution. These clock signals are particularly affected by technology scaling, in that long global interconnect lines become much more highly resistive as line dimensions are decreased. Increment in resistance introduces greater propagation delay along the interconnect lines.

Due to differences in interconnect delays on the clock distribution network, clock signals do not arrive at all of the flip-flops at the same time. Thus, there is a skew between the clock arrival times at different latches. Let X_i and X_j denote clock propagation delay from the clock-source to flip-flop FF_i and FF_j respectively. The clock skew is defined as $skew = X_i - X_j$. In synchronous digital systems, the circuit performance is directly proportional to its clock frequency. The clock skew and the logic delay between two adjacent sequential elements directly determine the lower bound of the clock period, or hence the upper bound of the clock frequency. This is seen from the following well-know inequality [1][2]:

$$\text{clock period} \geq t_d + t_{skew} + t_{su} + t_{ds}$$

where t_d is the delay on the longest path through combinational logic, t_{skew} is the clock skew, t_{su} is the set-up time of the synchronizing elements and t_{ds} is the propagation delay within the synchronizing element.

High-performance clock design is an area of active research. Previous researches focus on elimination of the clock skew ($t_{skew} = 0$) and reduce the terms t_{su} and t_{ds} with advances in VLSI fabrication technology, to maintain the minimization of clock period. Recently, some of the researchers treat the clock skew (t_{skew}) as a manageable resources and applying the $t_{skew} < 0$ to the synchronizing circuit, to further reduce the value of clock period. This negative value of clock skew ($t_{skew} < 0$) is known as useful-skew.

Clock routing is critical in high performance VLSI circuits design. Clock routing needs to meet skew constraint and circuit delay. Interconnect lines become highly resistive as line dimensions are decreased. This increased line resistance is one of the primary limitations for clock distribution networks on maximizing the performance of synchronous integrated circuit.

Due to nano-scale technology below 0.2um process, the interconnection delays contribute significantly to the signal propagation delay, and achieving an exact zero skew become more and more difficult. Finally, the control of any differences in the delay of the clock signals (or skew) can

severely limit the maximum performance of the entire system as well as create catastrophic race condition in which an incorrect data signal may latch within a register.

Clock signal is a major power consumer also. It switches at every clock cycle. The dynamic power due to the capacitance switching forms a dominant part of system power dissipation.

In this paper, a complete architecture of clock routing CAD sub-system for nano-scale VLSI design is proposed. (Figure 1) Previous researchers concentrate on developing their clock signal distribution technique in separate areas. Their efforts involve clock scheduling, clock signal routing with general skew constraint clock, initial clock tree topology generation, etc. We are first to introduce a complete clock routing CAD sub-system, implementing several clock signal distribution techniques introduced by previous researchers. The proposed sub-system manages to compute an optimized allowable clock period, and simultaneously synthesis a clock signal distribution routing path.

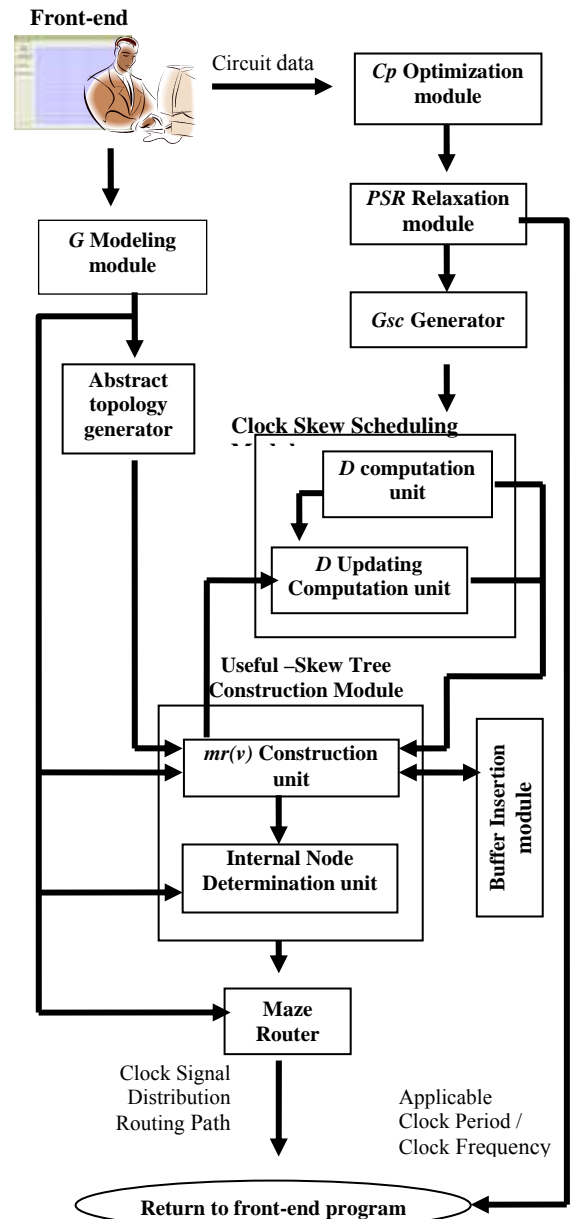
Due to limitation of the paper size, the scope of discussion in this paper focuses on useful-skew tree construction module only. The remainder of this paper is organized as follows. In Section II, the literature review of research work is presented. In Section III, the algorithm of useful-skew tree (UST) based on deferred-merge embedding (DME) paradigm is presented and in section IV, the proposed implementation work is shown.

2. Literature Review

High-performance clock design is an area of active research. Due to the ever-increasing die sizes and the continued scaling of devices and interconnects, the control of clock skew in a clock distribution network is rapidly becoming critical design problem. Previous work in this area can be divided into two categories: focusing either on clock skew optimization without considering layout synthesis, or on clock routing with simplified skew constraints.

The concept of scheduling the system-wide clock skews for improved performance while minimizing the likelihood of race condition was first presented by [3]. Fishburn presents a methodology in which a set of linear inequalities are solved using standard linear programming techniques in order to determine each clock signal path delay from the clock source to every clocked register. [4] also present an algorithm to optimize the clock cycles based on linear programming. [5] improves Fishburn's methodology based on the model given in Sakallah for determining an optimal clock schedule by selectively generating the shortest path constraints, permitting the inequalities describing the timing characteristics of each local data path to be solved more efficiently. [6] and [7] show that the clock skew problem can be solved using efficient graph-theoretic techniques in polynomial time. The idea of using graph methods is to take advantage of the structure of the problem to arrive at an efficient solution. [8] extend the Deoker's work to determine an optimal clock skew scheduling that is tolerant to process variations.

These techniques determine the skews among clock pins optimized, for example, performance and robustness. However, these approaches generated a clock schedule without considering its impact on the layout synthesis of the clock net.



Reference

Circuit data	The data of propagation delay of FF, combinational logic delay, and location of clock source / sinks.
C_p	Clock period / system frequency
PSR	Permissible Skew Range
G_{sc}	Skew Constraint Graph
D	All-pairs-shortest skew-constraint matrix
G	Graph Cartesian
Abstract topology	a binary tree such that all clock sinks are the leaf nodes of the binary tree.

Figure 1. Proposed Architecture of Clock Routing CAD Sub-system

The second category of works include H-tree method, which used regular systolic arrays [9], [10] and [11], Method of Means and Medians (MMM) algorithm proposed by [2], which generates a topology by recursively partitioning the set of sinks into two equal-sized subsets. [12] have proposed Geometric Matching Algorithm (GMA), a bottom-up matching approach to clock tree construction and [13] proposed the Weighted Center Algorithm (WCA). However, these methods focus primarily on path length balancing, rather than actual delay balancing.

Recently, the research work is focus on topology synthesis of the zero-skew tree (ZST) [14], and ZST routing based on the deferred-merge embedding (DME) algorithm [15], [16], [17], [18], [19], [20], [21] extended the DME algorithm to consider bounded-skew tree (BST) routing. Finally, [22] further enhance the BST algorithm and proposed an incremental scheduling technique to compute the feasible-skew range for constructing a useful-skew tree (UST) for general skew constraints.

3. Useful-Skew Tree (Ust) Algorithm Based on Deferred-Merge Embedding (Dme) Paradigm

The UST / DME routing technique consists of two main functions, which are Clock Skew Scheduling and Useful-Skew Tree Construction. (Figure 1) The Clock Skew Scheduling module consists of two computation unit, which is include the all-pairs-shortest skew-constraint matrix (D) computation unit and all-pairs-shortest skew-constraint matrix (D) updating computation unit. The all-pairs-shortest skew-constraint matrix (D) computation unit applies the Floyd-Warshall algorithm to compute the bounded limit of skew constraints for each sinks node (Figure 2). The all-pairs-shortest skew-constraint matrix (D) updating computation unit applies the incremental scheduling technique that proposed by [22] to narrow a nontrivial feasible skew constraint to a single skew value (Figure 3).

The Useful-Skew Tree Construction module also consists of two computation units. The first unit is merging region ($mr(v)$) construction unit, to perform the construction of a binary tree of merging region that represent the loci of possible embedding points of internal nodes in a bottom-up order (Figure 4). Secondly is internal node determination unit, which determines the exact location of internal nodes in a top-down order (Figure 5).

Input: original skew constraints matrix (W)
Output: All-pairs-shortest skew-constraints matrix (D)
<ol style="list-style-type: none"> 1. $n \leftarrow \text{rows}[W]$ 2. $D^{(0)} \leftarrow W$ 3. for $k \leftarrow 1$ to n 4. do for $i \leftarrow 1$ to n 5. do for $j \leftarrow 1$ to n 6. do $d_{ij}^{(k)} \leftarrow \min(d_{ij}^{(k-1)}, d_{ik}^{(k-1)} + d_{kj}^{(k-1)})$ 7. return $D^{(n)}$

Figure 2. Floyd-Warshall Algorithm

Input: skew commit $skew_{i,j} = x$, all-pairs shortest distance matrix, $D = \{d_{i,j}\}$
Output: an updated matrix D
<ol style="list-style-type: none"> 1. Set $d_{i,j} = -x$ 2. Set $d_{j,i} = x$ 3. for each $d_{k,l}$, $1 \leq k \neq l \leq n$ in D 4. Set $d_{k,l} = \min\{d_{k,l}, d_{k,i} - x + d_{j,l}, d_{k,j} + x + d_{i,l}\}$ 5. return matrix, D

Figure 3. Incremental Scheduling Algorithm

Input: clock pins / joining segment (Lu, Lw), local skew constraints
Output: merging region ($mr(v)$) or no solution
<ol style="list-style-type: none"> 1. Determine boundary of shortest-distance region (SDR) among Lu, Lw 2. for $n = Lu$ to Lw along the boundary of SDR 3. delay (t_u) = Elmore Delay (n, Lu) 4. delay (t_w) = Elmore Delay (n, Lw) 5. clock skew ($Tskew$) = $t_u - t_w$ 6. if $Tskew \in FSR_{i,j}$ 7. mark (each n at same Manhattan arc and $\in SDR$) \in merging region ($mr(v)$) 8. return $mr(v)$

Figure 4. Bottom-up Phase Merging Region Construction Algorithm

Input: Tree of merging region $Tmr(v)$
Output: Location of internal nodes
<ol style="list-style-type: none"> 1. for each internal node (v) in $Tmr(v)$ 2. if v is the root 3. Choose any node ($l(v)$) \in $mr(v)$. 4. else 5. Let P be the parent node of v 6. Let Q be the merging region of v's sibling node 7. if e_v not determined during merging region construction 8. $e_v = \text{shortest distance of joining segment } Q \text{ and node } l(p)$ 9. $= d(JS_Q(mr(v)), l(p))$ 10. Choose any $l(v) \in JS_Q(mr(v))$ closet to $l(p)$

Figure 5. Top-down Phase Internal Node Determination Algorithm

4. Implementation Of Ust/Dme Algorithm

The functional block diagram of proposed useful-skew tree construction module is provided in Figure 6. The proposed useful-skew tree construction module is targeted to be implemented in C++. The module will receive the data of abstract topology, location of clock source and clock sinks, and each local skew constraint as an input; to produce the useful-skew tree as an output that include the

location of each internal node and the linking relationship information to their child nodes.

Referring to the diagram, the proposed useful-skew tree construction module consists of three main design components, which is briefly described as follows:

- **Computation units:** These computation units include the D computation unit, D Updating Computation unit, $mr(v)$ Construction unit, Joining Segment Determination unit, delay modeling unit and Internal Node Determination unit. The computation units will perform the algorithms described in the previous section. The bottom-up phase merging region construction algorithm (Figure 4) is simultaneously performed by $mr(v)$ construction unit, joining segment determination unit and delay modeling unit.
- **Control units:** The control unit handles the data transfer (in/out) among the computation units, and data transfer to/from routing database. The control unit also controls the whole processing flow during the useful-skew tree construction.
- **Routing database:** The database is used to store the information such as location of each clock sinks, location of internal nodes, joining segments, all merging regions, linking relationship information and some necessary identifiers that will be used for trace back process.

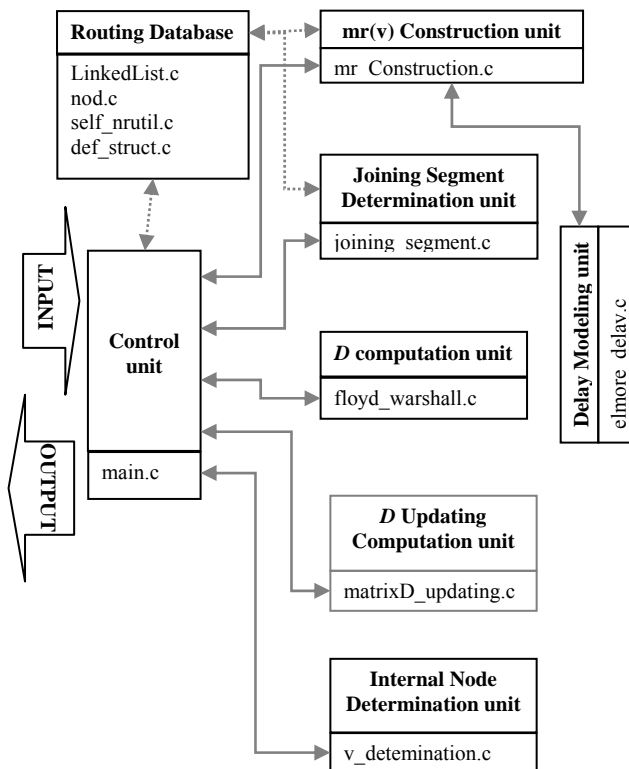


Figure 6. Functional Block Diagram of proposed useful-skew tree construction module

5. Conclusion

This paper has proposed the design of useful-skew tree construction module to synthesize the useful-skew tree routing path that implemented UST/DME algorithm. The advantage of this proposed architecture is that the clock layout is constructed and simultaneously determines the skew schedule, such that the resulting skew schedule is not only feasible, but also best for routing in term of wire-length.

References

- [1] H. Bakoglu, *Circuits, Interconnections and Packaging for VLSI*. Reading, MA: Addison-Wesley, 1990.
- [2] M. A. B. Jackson, A. Srinivasan, and E. S. Kuh, *Clock routing for high performance IC's*, In Proc. ACM / IEEE Desing Automation Conference, pp. 573-579, 1990.
- [3] J. P. Fishburn, *Clock skew optimization*. IEEE Trans. on Comput. 39, 7 (July), 945-951. 1990.
- [4] K. A. Sakallah, T. N. Mudge, and O. A. Olukotun, *checkTc and minTc: Timing verification and optimal clocking of synchronous digital circuits*. In Proc. International Conference on Computer Aided Design, 552-555. 1990.
- [5] T. Szymanski, *Computing optimal clock schedules*. In Proc. Design Automation Conference, 399-404. 1992.
- [6] R. B. Deokar, and S. S. Sapatnekar, *A graph-theoretic approach to clock skew optimization*. In Proc. IEEE International Symposium on Circuits and Systems, 407-410. 1994
- [7] S. S. Sapatnekar, and R. B. Deokar, *Utilizing the retiming-skew equivalence in practical algorithm for retiming large circuit*. IEEE Trans. Computer Aided Design Integrated Circuit System 15, 1237-1248. 1996.
- [8] J. L. Neves, and E. G. Friedman, *Optimal clock skew scheduling tolerant to process variations*. In Proc. Design Automation Conference, 623-628. 1996.
- [9] H. Bakoglu, J.T. Walker, and J.D. Meindl, *A symmetric clock-distribution tree and optimal high-speed interconnection for reduced clock skew in ULSI and WSI circuit*. In Proc. IEEE International Conference on Computer Design, pp. 118-122. 1986.
- [10] D. Dhar, M.A Franklin, and D.F. Wann, *Reduction of clock delays in VLSI structure*. In Proc. IEEE International Conference on Computer Design, pp. 778-783. 1984.

- [11] A.L. Fisher, and H.T. Kung, *Synchronizing large systolic arrays*. Proc. of SPIE, pp. 44-52. 1982.
- [12] A. B. Kahng, J. Cong, and G. Robins, *Matching based Models for high performance clock routing*. IEEE Transaction on CAD of Integrated Circuit and System, 12:1157-1169. 1993.
- [13] N.A. Sherwani, and B. Wu, *Clock layout of high performance circuits based on weighted center algorithm*. In Proc. Fourth IEEE International ASIC Conference and Exhibit, pages p15-5.1-5.4. 1991.
- [14] R.S. Tsay, *An exact zero-skew clock routing algorithm*. IEEE Transaction Computer Aided Design Integrated Circuit System CAD-12, 242-249. 1993.
- [15] M. Edahiro, *Minimum path-length equi-distant routing*. In Proc. IEEE Asia-Pacific Conference on Circuits and Systems (December), 41-46. 1992.
- [16] T.H. Chao, Y.C. H. Hsu, J.M. Ho, K. D. Boese, and A. B. Kahng, *Zero skew clock routing with minimum wirelength*. IEEE Transaction Circuit System 39, 799-814. 1992.
- [17] A. B. Kahng, and C. W. A. Tsao, *Planer-DME: A single -layer zero-skew clock tree router*. IEEE Transaction Computer Aided Design Integrated Circuit System 15, 8-19. 1996.
- [18] J. Cong, and C.K. Koh, *Minimum-cost bounded-skew clock routing*. In Proc. IEEE International Symposium on Circuit and System, 1.215-1.218. 1995.
- [19] J.H. Huang, A. B. Kahng, and C. W. A. Tsao, *On the bounded-skew routing tree problem*. In Proc. Design Automation Conference (June), 508-513. 1995.
- [20] A. B. Kahng, and C.W. A. Tsao, *Practical bounded-skew clock routing*. J. VLSI Sig. Process. (Special Issue on High Performance Clock Distribution Networks) 16, 199-215. 1997.
- [21] J. Cong, A. B. Kahng, C.K. Koh, and C.W. A. Tsao, *Bounded-skew clock and Steiner routing*. ACM Transaction Design Automation Electronic System 3, 3, 341-388. 1998.
- [22] C. W. A. Tsao, and C. K. Koh, *UST/DME: A clock tree router for General Skew Constraints*. ACM Transactions on Design Automation of Electronic System, Vol. 7, No. 3. 2002.

Analysis of Throughput S-ALOHA CDMA with Differential MRC at Two Models Fading Channels

Hoga Saragih, Gunawan Wibisono and Eko Tjipto Rahardjo

Department of Electrical Engineering, Faculty of Engineering,
The University of Indonesia, Kampus Baru UI Depok, 16424, Jakarta, Indonesia
E-mail: hoga15876@yahoo.com, gunawan@eng.ui.ac.id and eko@eng.ui.ac.id

Abstract

Fading and interference are factors that limit wireless communication system. Fading can reduce the quality of received signal, meanwhile Interference is caused by frequency reuse in cellular communication system. In cellular system with co channel interference, desired signal and interference signal could have different characteristics. Diversity technique of maximal ratio combining (MRC) is recommended to overcome co channel interference by modeling channel condition as Nakagami/Nakagami fading channel with diversity technique of L branch antenna, and also modeling fading channel condition as Nakagami/Nakagami fading channel and Rician/Nakagami fading channel.

Bit error rate (BER) is one of the parameters to identify CDMA system performance, by using Gaussian approximation method and exact solution. By using both methods, close form equation for BER can be derived with any condition of fading, detection scheme, and diversity technique. BER equation is derived by the Gaussian approximation and exact solution in Nakagami/Nakagami fading channel and Rician/Nakagami fading channel.

MRC diversity technique with L branch differential detection is used to measure throughput of S-ALOHA CDMA. By using two different fading models between desired signal and interference signal, throughput can be analyzed by Nakagami/Nakagami and Rician/Nakagami channel condition.

MRC diversity and differential detection (DPSK) with L branch antenna exist to handle multipath fading, interference and to overcome the system performance. The result indicates that the deployment of L branch antenna on the receiver will better system throughput. The more interference signals, the worse throughput system. The higher Nakagami parameter, m , and Rician factor, K , of the desired signal and interference signal the better throughput.

Keywords: S-ALOHA CDMA, BER, MRC, Nakagami, and Rician.

1. Introduction

In wireless communication mostly damaged package due to multipath fading and interference. Fading can degrade the quality of received signal and interference due to the use of frequency reuse in cellular communication system, Co Channel Interference (CCI) is one of the interference resulted from the use of frequency reuse. At cellular system affected by CCI, desired signal and interference signal can have different fading characteristic.

Performance analysis of ALOHA with fading channel has been done by A. Sheikh *et al* [1]. Performance analysis of S-ALOHA in Rayleigh fading channel has been done by J.A. Roberts *et al* [2] and J. C. Anbark *et al* [3]. And also the performance of S-ALOHA Nakagami/Nakagami fading channel with same parameter value for the test packet and interfering packets has also done by A. Al-Semari *et al* [4]. Besides those various approaches above, calculating various channel model of fading, which is distributed fading channel

of Rayleigh, Rician and Nakagami, has done calculation of BER CDMA system.

Abdel Hafez *et al* [5] analyzed the performance of receiver uplink/downlink CDMA with MRC diversities technique with coherent detection and differential detection and influenced by existence of multiple access interference (MAI). In [6], desired signal and interference signal have the condition of Rayleigh fading. Then, [7] modeling desired signal to have the Rician fading condition and signal interference to have the Rayleigh fading condition.

To overcome CCI Y. D. Yao *et al* suggested the usage of technique selection combining diversities and model condition of Nakagami/Nakagami fading channel [8]. Damar suggested the usage of MRC diversities technique but still having one antenna and models condition of Nakagami/Nakagami fading channel [9]. Gunawan suggested the MRC diversities and models condition as Nakagami/Nakagami and Rician/Nakagami fading channel with the usage of diversities technique with L branch antenna [10].

This paper is suggested the usage of technique CDMA ALOHA by analyzing throughput CDMA ALOHA with differentials MRC at two models Nakagami/Nakagami and Rician/Nakagami fading channel.

Part 2 of this paper explains model system, part 3 explains calculating and analysis the result, and part 4 represents conclusion.

2. Model System

2.1. Transmitter Model

Each user transmitted power average P_k , ($1 \leq k \leq K$), at

$f_c = \frac{\omega_c}{2\pi}$ frequency carrier, with $R_b = \frac{1}{T_b}$ data rate and

$R_c = \frac{1}{T_c}$ chip rate. Each user has a unique spreading code

sequence unique, $a_{k,j} \in \{ -1, +1 \}$ with the following code waveform:

$$a_k(t) = \sum_{j=-\infty}^{\infty} a_{k,j} p_{T_c}(t - jT_c) \quad (1)$$

Signal data user k , with elements data $b_{k,j} \in \{ -1, +1 \}$ expressed:

$$b_k(t) = \sum_{j=-\infty}^{\infty} b_{k,j} p_{T_b}(t - jT_b) \quad (2)$$

Where $p_T(t)$ is rectangular pulse unit (T duration), T_c is duration chip, T_b is duration bit, $\eta = \frac{T_b}{T_c}$ is processing gain.

Transmitted signal k by user is:

$$S_k(t) = \sqrt{2P_k} a_k(t) b_k(t) \cos(\omega_c t + \Phi_k) \quad (3)$$

Where the phase angle Φ_k assumed distributed uniformly $[0, 2\pi]$. $S_T(t)$ is total signal all signals of k user of system, expressed as follows:

$$S_T(t) = \sum_{k=1}^K S_k(t) \quad (4)$$

2.2. Channel Model

In the channel model, a linear filter models fading type where the users k characterized by a complex-valued low pass equivalent impulse response as follow:

$$h_k(t) = \sum_{l=0}^{L_k-1} \beta_k^l e^{j\theta_k^l} \delta(t - \tau_k^l) \quad (5)$$

Where β_k^l is the attenuation factor, τ_k^l is delay, θ_k^l is phase, L_k is the channel of the user to k , $\delta(\cdot)$ is direct delta function.

2.3. Receiver Model

Fig. 1 shows the received signal, $r(t)$ of output channel mobile is time convolution $S(T)$ and $h(t)$ which is expressed as follows:

$$r(t) = \sum_{k=1}^K \sum_{l=0}^{L_k-1} \beta_k^l \sqrt{2P_k} a_k(t - \tau_k^l) b_k(t - \tau_k^l) \cos(\omega_c t + \tau_k^l) + n(t) \quad (6)$$

where $n(t)$ is AWGN with zero mean and two sided powers spectral densities $N_0/2$.

If in MRC uses L branch antenna and has interference co channel signal I , therefore total output SINR of MRC can be expressed [11] as:

$$r = \frac{\sum_{p=1}^L \Omega_j^p}{I + \sum_{q=1}^L \Omega_t^q} = \frac{X}{I + Y} = \frac{X}{\omega} \quad (7)$$

Where Ω_j^p shows average SNR of desired signal j of antenna branch p , Ω_t^q shows average INR of interference signal t of antenna branch q , $X = \sum_{p=1}^L \Omega_j^p$ shows total

SNR desired signal, and $Y = \sum_{q=1}^L \Omega_t^q$ shows total INR interference signal each antenna branch.

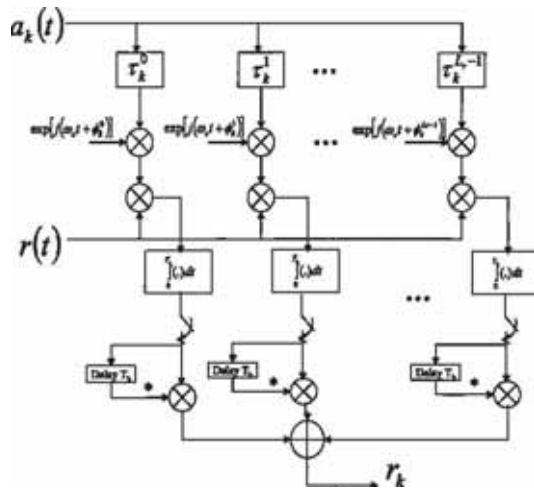


Fig.1. Receiver using MRC with differential detection [5]

PDF SINR at MRC shown by [11]:

$$f_r(r) = \int_0^{\infty} (\omega + I) f_X((\omega + I)r) f_Y(\omega) d\omega \quad (8)$$

2.4. BER with Two Fading Models

To measure the system performance, the equation of bit error rate (BER) will be applied by using two fading models. These two fading models are desired signal and interference

signal with different fading condition. If $X = \sum_{p=1}^L \Omega_j^p$ is total SNR of desired signal, Ω_j is average SNR per antenna branch and $\Gamma(\cdot)$ is gamma function. So distributed PDF of desired signal of Rayleigh described [12] as follows:

$$f_X(x) = \frac{x^{L-1}}{(\Omega_j)^L \Gamma(L)} e^{-\frac{x}{\Omega_j}} \quad (9)$$

If m_j is the Nakagami parameter of desired signal, distributed PDF of desired signal of Nakagami described [12] as follows:

$$f_X(x) = \frac{m_j^{m_j L}}{\Omega_j^{m_j L}} \frac{x^{m_j L - 1}}{\Gamma(m_j L)} e^{-\frac{m_j x}{\Omega_j}} \quad (10)$$

If K_j is Rician factor desired signal and $I_n(\cdot)$ is modified Bessel function order to n , distributed PDF of desired signal of Rician described [12] as follow:

$$f_X(x) = \frac{(K_j + 1)}{\Omega_j} \frac{x(K_j + 1)^{\frac{L-1}{2}}}{K_j L \Omega_j} e^{-K_j L \frac{(K_j + 1)}{\Omega_j}} I_{L-1} \left(2 \sqrt{\frac{K_j L (K_j + 1)}{\Omega_j}} x \right) \quad (11)$$

If $Y = \sum_{q=1}^I \Omega_t^q$ is total INR interference signal, Ω_t is average INR of interference signal, and $\Gamma(\cdot)$ is gamma function, so distributed PDF of interference signal of Rayleigh described [12] as follows:

$$f_Y(y) = \frac{y^{I-1}}{(\Omega_t)^I \Gamma(I)} e^{-\frac{y}{\Omega_t}} \quad (12)$$

If m_t is Nakagami parameter interference signal, distributed PDF of interference signal of Nakagami described [12] as follows:

$$f_Y(y) = \frac{m_t^{m_t I}}{\Omega_t^{m_t I}} \frac{y^{m_t I - 1}}{\Gamma(m_t I)} e^{-\frac{m_t y}{\Omega_t}} \quad (13)$$

If K_t is Rician factor interference signal, and $I_n(\cdot)$ is Bessel function modified order to n , distributed PDF of interference signal of Rician described [12] as follows:

$$f_Y(y) = \frac{(K_t + 1)}{\Omega_t} \frac{y(K_t + 1)^{\frac{I-1}{2}}}{K_t I \Omega_t} e^{-K_t I \frac{(K_t + 1)}{\Omega_t}} I_{I-1} \left(2 \sqrt{\frac{K_t I (K_t + 1)}{\Omega_t}} y \right) \quad (14)$$

General equation to calculate bit error rate (BER) determine by [5]:

$$P_e = \int_0^\infty f_r(r) P(r) dr \quad (15)$$

Where $f_r(r)$ is PDF SINR, $P(r)$ is bit error probability from scheme detected at AWGN channel. If receiver uses differential detection (DPSK) in AWGN channel, therefore:

$$P(r) = \frac{1}{2} e^{-r} \sum_{k=0}^{L-1} b_k r^k \quad (16)$$

where:

$$b_k = \frac{1}{k!} \sum_{j=0}^{L-1-k} \binom{2L-1}{j} \quad (17)$$

2.4.1. Rayleigh/Rayleigh

To calculate BER model Rayleigh/Rayleigh, first we determined the PDF of SINR for Rayleigh/Rayleigh by substituting (9) and (12) to (8) and (16) the equation will become:

$$P_e = \frac{1}{(\Omega_j)^L \Gamma(L) \Gamma(I)} \frac{1}{2} \sum_{k=0}^{L-1} b_k \Gamma(L+k) \sum_{i=1}^N w_i (\Omega_j x_i + I) x_i^{L-1-k} \frac{(\Omega_t x_i + I)^{L-k}}{\Omega_j} + I \quad (18)$$

2.4.2. Rayleigh/Nakagami

PDF from SINR with the condition of Rayleigh/Nakagami can be determined by applying the equation (9) and (13) to the equation (8) and (16), so the equation obtained [13], [14]:

$$P_e = \frac{1}{(\Omega_j)^L \Gamma(L) \Gamma(m_t I)} \frac{1}{2} \sum_{k=0}^{L-1} b_k \Gamma(L+k) \sum_{i=1}^N w_i \frac{\Omega_t}{m_t} x_i + I x_i^{m_t I - 1 - k} \frac{\Omega_t}{\Omega_j} x_i + I \quad (19)$$

Equation (19) is the BER system of Rayleigh/Nakagami fading channel. Further more, if $m_t = 1$ it will reduce to be Rayleigh/Rayleigh equation (18).

2.4.3. Nakagami/Rayleigh

PDF from SINR with condition of Nakagami/Rayleigh can be determined by applying the equation (10) and (12) to equation (8) and (16), so the equation obtained is shown below:

$$P_e = \frac{m_j}{\Omega_j} \frac{1}{\Gamma(m_j L) \Gamma(L)} \frac{1}{2} \sum_{k=0}^{L-1} b_k \Gamma(m_j L + k) \sum_{i=1}^N w_i (\Omega_t x_i + 1)^{m_j L} x_i^{L-1} \frac{m_j (\Omega_t x_i + 1)}{\Omega_j} + 1 \quad (20)$$

Equation (20) is BER system of channel Nakagami/Rayleigh fading channel.

2.4.4. Nakagami/Nakagami

PDF from SINR with condition of Nakagami/Nakagami can be determined by applying the equations (10) and (13) to equation (8) and (16), so the equation obtained is shown below:

$$P_e = \frac{m_j}{\Omega_j} \frac{1}{\Gamma(m_j L) \Gamma(m_t L)} \frac{1}{2} \sum_{k=0}^{L-1} b_k \Gamma(m_j L + k) \sum_{i=1}^N w_i x_i^{m_t L} \frac{m_j \frac{\Omega_t}{m_t} x_i + 1}{\Omega_j} + 1 \quad (21)$$

The equation (21) is BER system of Nakagami/Nakagami fading channel.

2.4.5. Rician/Rayleigh

PDF from SINR of the condition of Rician/Rayleigh can be determined by applying the equation (11) and (12) to equation (8) and (16), so the equation obtained is shown below:

$$P_e = \frac{(K_j + 1)^L}{\Omega_j} e^{K_j L} \frac{1}{\Gamma(L) \Gamma(L)} \frac{1}{2} \sum_{k=0}^{L-1} b_k \Gamma(L + k) \sum_{i=1}^N w_i (\Omega_t x_i + 1)^L x_i^{L-1} \frac{(K_j + 1)(\Omega_t x_i + 1)}{\Omega_j} + 1 \quad (22)$$

Where ${}_1F_1(a, b, z)$ is Kummer confluent hyper geometric function, the equation (22) above is BER Rician/Rayleigh, if $K_j = 0$ it will reduce to be Rayleigh/Rayleigh (18).

2.4.6. Rician/Nakagami

PDF from SINR with Rician/Nakagami condition can be determined by applying the equation (11) and (13) to the equation (8) and (16), so the equation obtained is shown below:

$$P_e = \frac{(K_j + 1)^L}{\Omega_j} \frac{1}{\Gamma(m_t L) \Gamma(L)} e^{K_j L} \frac{1}{2} \sum_{k=0}^{L-1} b_k \Gamma(L + k) \sum_{i=1}^N w_i \frac{\Omega_t}{m_t} x_i + 1 \frac{(K_j + 1) \frac{\Omega_t}{m_t} x_i + 1}{\Omega_j} + 1 \quad (23)$$

If $K_j = 0$, and $m_t = 1$ it will reduce to be Rayleigh/Rayleigh in equation (18).

2.5. Throughput S-ALOHA CDMA

Throughput S-ALOHA CDMA in fading channel models given by [3], [4], and [15]:

$$S = G \sum_{n=1}^{\infty} P_n(n) (1 - P_e)^L \quad (24)$$

where L is packet length and $P(n)$ is the probability of n frame obtained from certain frame time of S-ALOHA given by the distribution Poisson:

$$P_n(n) = \frac{G^n e^{-G}}{n!} \quad (25)$$

3. Calculating and Analyzing

Fig. 2 shows throughput of S-ALOHA CDMA in Rayleigh/Rayleigh fading channel by varying total antenna of receiver ($L = 1, 2, 4, 6, 8, 10$). Calculation is done for total interference signal, $I = 3$, and average INR = 1 dB. From Fig.

2 shows that S-ALOHA CDMA throughput will better by increasing total antenna of receiver (L). The throughput increases due to the smaller of noise effect and interference so that the throughput becomes better. Total available antenna very influences system the performance. The more available antenna is the more desired signal can be captured. Those signals then combined with diversity technique of MRC therefore optimal throughput can be obtained.

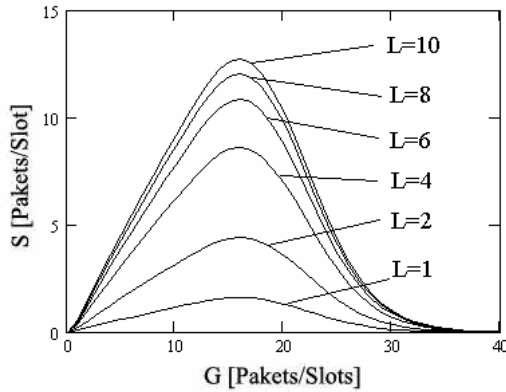


Fig. 2. Throughput of S-ALOHA CDMA in Rayleigh/Rayleigh fading channel with parameter average SNR = 10 dB and average INR = 2 dB.

Fig. 3 shows throughput of S-ALOHA CDMA in Nakagami/Rayleigh fading channel with varied total antenna receiver. The calculation is using Nakagami- m parameter = 2, average SNR = 4 dB and average INR = 1 dB. Fig. 3 shows the effect of total L against system throughput, Fig. 3 is similar with Fig. 2 but in Fig. 3 fading model Nakagami is described as desired signal, and Rayleigh is described as interference signal.

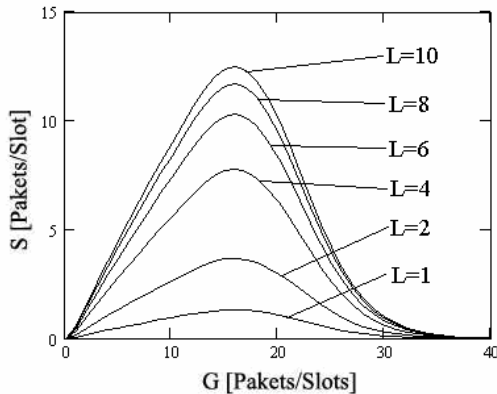


Fig. 3. Throughput of S-ALOHA CDMA in Nakagami/Rayleigh fading channel with parameter $m_j = 2$, average SNR = 4 dB and average INR = 1 dB

Fig. 4 shows throughput of S-ALOHA CDMA in Rayleigh/Nakagami fading channel with varied number of receiver antenna. The calculation is using Nakagami- m parameter = 2, average SNR = 4 dB and average INR = 1 dB. Fig. 4 shows the effect of total L against throughput system. The bigger L the more increases throughput as total interference signal and the change of throughput of Rayleigh/Nakagami are not too big because Nakagami is

described as interference signal while Rayleigh is described as desired signal.

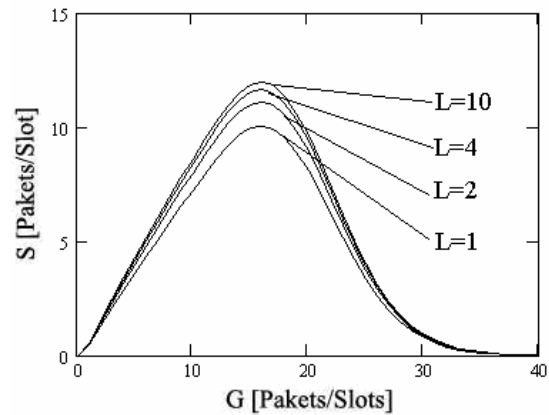


Fig. 4. Throughput of S-ALOHA CDMA Rayleigh/Nakagami fading channel with parameter average INR = 2 dB, $I = 2$ and $L = 6$

Fig. 5 shows throughput of S-ALOHA CDMA in Nakagami/Nakagami fading channel with varied parameter Nakagami- m interference signal (m_i). The figure shows that the bigger value of Nakagami- m , the better the throughput. Such condition is related to the theory that the bigger the value of Nakagami- m parameter, fading condition is getting smaller. If the value of (m_i) is became bigger to be unlimited, the channel condition will be considered as non-fading channel.

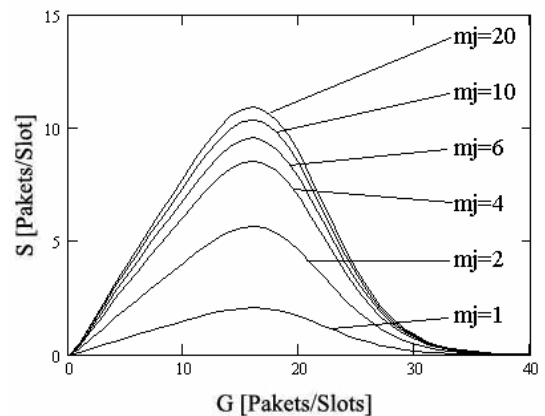


Fig. 5. Throughput of S-ALOHA CDMA in Nakagami/Nakagami fading channel with parameter INR = 1 dB, SNR= 10 dB, $m_i = 2$ and $L = 2$.

Fig. 6 shows throughput of S-ALOHA CDMA in Rician/Nakagami fading channel. The calculation using parameter of average INR = 2 dB, total antenna of receiver, $L = 6$, total interference signal $I = 2$, and parameter of Nakagami- m interference signal (m_i) = 2. Fig. 6 also shows the effect of fading against throughput system. The bigger Rician factor K_j the bigger total signal of LOS, so the throughput is getting better. For big value of K_j , the

difference inter graph is not too big. This is because value of K_j is getting higher, so fading is getting smaller. If the value of K_j is enlarged to the unlimited the channel condition is therefore considered as non-fading. For this condition, the increasing of Rician factor K_j has no effect at all against the system performance.

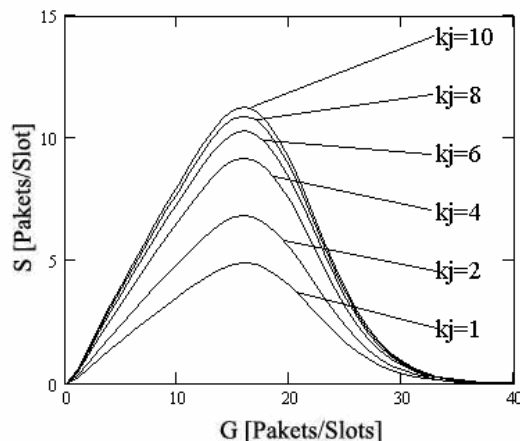


Fig. 6. Throughput of S-ALOHA CDMA in Rician/Nakagami fading channel with parameter average $\text{INR} = 2$ dB, $L = 6$, $I = 2$ and $(m_t) = 2$.

4. Conclusion

Analysis of throughput S-ALOHA CDMA performance with MRC diversities and differential detection has been analyzed at two fading model. Those channel models are Nakagami/Nakagami and Rician/Nakagami. Throughput system becomes better if total available antenna of receiver is getting bigger. The bigger fading parameter (Nakagami and Rician), the better throughput system.

References

- [1] A. Sheikh, Y-D Yao and X. WU, "The ALOHA System in Shadowed Mobile Radio Channel with Slow or Fast Fading", IEEE Transactions on Vehicular Technology, Vol.39, No.4, November 1990, pp.289-286.
- [2] J. A. Roberts and T. J Healy, "Packet Radio Performance Over Slow Rayleigh Fading Channel", IEEE Transaction on Vehicular Technology, Vol. COM-28, No.2, 1980, pp.279-286.
- [3] J. C. Anbark and W. Van Blitterswijk, "Capacity of Slotted ALOHA in Rayleigh Fading Channel", IEEE Journal on Selected Areas in Communication, Vol. SAC-5, No.2, February 1987, pp.261-265.
- [4] Sand A. Al-Semari and Mohsen Guizani, "Channel Throughput of Slotted ALOHA in Nakagami Fading Environment", IEEE Transaction on Communications, 1997.
- [5] M. Abdel Hafez and F. Alagoz, "A Generalized Performance Study of DS-CDMA Uplink/Downlink Receivers in Nakagami Wideband Fading Channel," IEICE Trans. Commun., vol. E88-B, no.1, pp. 333-344, January 2005.
- [6] K. W. Sowerby and A. G. Williamson, "Outage Probability Calculations for A Mobile Radio System Having Multiple Rayleigh Interferes," Electron. Lett., vol.23, pp. 600-601, 1987.
- [7] Y. D. Yao and A. U. H. Sheik, "Outage Probability Analysis for Micro cell Mobile Radio Systems with Co channel Interferers in Rician/Rayleigh Fading Environment," Electron. Lett., vol.26, pp. 864-866, June 1990.
- [8] Y. D. Yao and A. U. H. Sheik, "Investigations into Co channel Interference in Microcellular Mobile Radio Systems," IEEE Trans. on Vehicular Technology., vol.41, no.2, May 1992.
- [9] Damar Wijaya and Gunawan Wibisono, "Analysis BER CDMA with Two Model Path Gain Channel Fading Nakagami," Proc. The 6 th Seminar on Intellegent Technology and Its Applications (SITIA 2005), Institut Teknologi Sepuluh November, Surabaya, 2 May 2005.
- [10] Gunawan Wibisono and Andy Prakoso, "Performance Analysis CDMA with Differential MRC at two Model Fading," Proc. EECCIS 2006, Brawijaya University Malang, Part 1, C-135, May 16, 2006.
- [11] Chirasil Chayawan and Valentine A. Aalo, "Average Symbol Error Rate of Maximal Ratio Combining Scheme in the Multiple Co channel Interferers", ECTI Trans. on Electrical Eng., Electronics, and Communications, Vol.1, No.1, Augustus 2003.
- [12] Hong-Cuan Yang and Mohammed-Slim Alouini, "Closed-form Formulas for the Outage Probability of Wireless Communication System with a Minimum Signal Power Constraint", University of Minnesota, 16 October 2002.
- [13] I. Gradshteyn and I. Ryzhik, *Tables of Integrals, series and product*, academic, New York, 1980.
- [14] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, 1970.
- [15] M. S. Do, Y. J. Park, J. Y. Lee, "The Effect of Spreading Gain Control on a CDMA Slotted ALOHA System," IEEE Trans, Computer Commun. Vol. 26, pp 996-1006, July 20, 2002.

Application of Fuzzy Logic in Active Vibration Control of Beam

Amirasyid D^{*} ¹, Momoh Jimoh E.S²

¹ Department of Mechanical Engineering, Faculty of Engineering
Universiti Tenaga Nasional, 43009 Kajang Selangor Malaysia
Tel: +60-31-89212020 Ext 6330, Fax: +60-3-89212116, E-mail: amirasyid@uniten.edu.my

² Department of Mechatronic Engineering, Kulliyah of Engineering
International Islamic Universiti Malaysia, P.O Box 10, 50728 Kuala Lumpur Malaysia
Tel: +60-3-61964494, Fax: +60-3-20564466, E-mail: Momoh@iiu.edu.my

Abstract

This paper presents a real-time implementation of fuzzy logic in active vibration control of cantilever beam bonded with piezoelectric patches. Direct fuzzy algorithm is developed to control the first natural mode of vibration of beam. The control performance of fuzzy logic is compared with positive position feedback, a widely used algorithm for active vibration control. To further investigate the ability of fuzzy, mass is added to the beam to introduce the uncertainty in the model parameter. The real-time-windows-target (RTWT) under the SIMULINK /MATLAB environment is utilized for the real time implementation of the system. The experimental results show the effectiveness of fuzzy algorithm to suppress the vibration even in the presence of uncertain model parameter.

Keywords: Active vibration control, Fuzzy logic.

1. Introduction

Fuzzy logic theory has been studied in active vibration control (AVC) as an alternative to the classical as well as modern control theories. The main advantage of fuzzy logic (FL) that attracts great interest of researchers is its ability to suppress the vibration without having to know the exact mathematical model of the system to be controlled. Apart from that, it is also capable of modeling the nonlinearities in the system qualitatively and hence eliminating the problem of unmodelled dynamics and uncertainties in model parameters. These features are useful when dealing with real world cases where systems are complex, nonlinear and hence difficult to model mathematically.

Among the earliest work that utilized FL in AVC was by Tsoukkas and Vanlandingham [1]. Two inputs and seven membership functions were used that resulted in 49 fuzzy rules. The effectiveness of fuzzy was demonstrated theoretically using beam model. However a large number of rules pose a difficulty in real-time hardware implementation of fuzzy algorithm since all the rules must be processed in parallel. Kwak and Sciulli [2] proposed a reduction of rules by utilizing the variable structure system (VSS) with a straight switching line. Significant reductions of rules were demonstrated from 49 to only four rules. Cantilevered beam bonded with piezoelectric sensors and actuators was used in real time implementation of the system.

Sharma et al, [3] used fuzzy logic with modal space control to suppress the vibration of beam. Nine fuzzy rules based on modal displacement and velocities were used in real time implementation of the system. Effectiveness of fuzzy was demonstrated by comparing the results with independent modal space control (IMSC) and modified independent modal space control (MIMSC). Mayhan and Washington [4] proposed improvement in fuzzy by incorporating learning ability into fuzzy. The algorithm was named as fuzzy model reference learning control. In this study, membership functions were allowed to change depending upon the learning mechanism. The experimental results showed the improvement in settling time of this approach compare to direct fuzzy.

Several other studies proposed the use of fuzzy logic in conjunction with other control approaches. Gu and Song [5] utilized fuzzy to adjust the scalar gain of Positive Position Feedback (PPF) controller. The authors demonstrated the successful application of fuzzy to reduce the initial overshoot of PPF controller. Zeinoun and Khorrami [6] proposed fuzzy system to incorporate the adaptive capability on the control algorithm. The authors used fuzzy to co-ordinate between three controllers developed based on linear quadratic regulator with frequency-dependent weightings operated at three different operating regimes.

In this study, the application of FL in its pure form is investigated using cantilevered beam bonded with piezoelectric patches. The number rules and its impact on the performance of AVC are investigated. The real-time implementation of fuzzy is carried out using

^{*} Corresponding Author. E-mail: amirasyid@uniten.edu.my
Tel: +60-3-892212020, Fax: +60-3-89212116

Real-Time-Windows-Target (RTWT) toolbox in Matlab/Simulink environment. All the results are compared with PPF controller, a widely used active control algorithm in order to measure the effectiveness of FL. To further study the effectiveness of FL, mass is added into the tip of the beam to introduce the uncertainty in the model parameter. The ability of FL to cope with the uncertainty is then evaluated.

2. Active Vibration Control

Vibration control can be broadly classified into two categories namely passive control and active control. In passive control, vibration is eliminated by manipulating either the mass, damping or stiffness of the system. Mass is usually difficult to adjust but the damping and the stiffness can be varied to reduce the vibration level. In active control, on the other hand, vibration energy is dissipated through the external power applied to the actuator to counteract with the vibration force. AVC can be regarded as destructive interference whereby the secondary wave generated from the actuator interact and cancel/reduce the unwanted vibration on the system. The main part in active control is the control algorithm that determines the required force to be generated by the actuator in order to reduce or eliminate the vibration. The following briefly explains each control algorithm used in this study.

2.1. Fuzzy Logic Controller (FLC)

FLC is the controller that uses the expert knowledge in the form of linguistic rules to determine the control strategy. The linguistic rules are constructed based on the observable phenomenon and human knowledge on the system. FL does not require mathematical model of the system. The basic configuration of FLC contains three main components which are the fuzzification, inference engine and defuzzification. Figure 1 shows the architecture of FLC. For details explanation, readers can refer to the book by Passino and Yurkovich in reference [7].

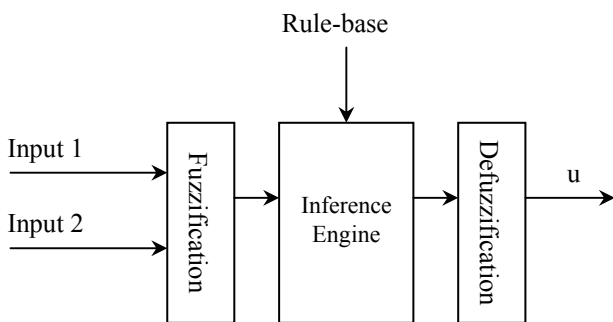


Figure 1: Achitecture of FLC

2.2 Positive Position Feedback Controller

PPF controller is widely used algorithm in active vibration control of structure. It was first proposed by Goh and Caughey [8]. Several works on PPF in AVC can be found in [5-6, 9]. The basic principle of PPF is that the structural position information is fed to the compensator in

the feedback loop that has the following transfer function

$$\frac{K\omega_c^2}{s^2 + 2\zeta_c\omega_c s + \omega_c^2} \tag{1}$$

Where ω_c and ζ_c are natural frequency and damping ratio of the controller respectively. K is the scalar feedback gain that tunes the controller to provide the best response. The block diagram of PPF controller to control the beam system is shown in figure 2.

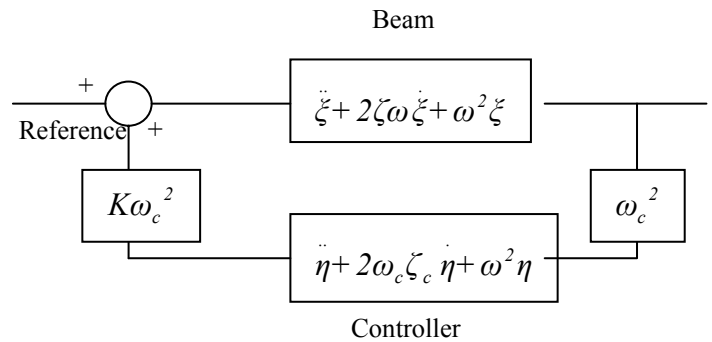


Figure 2: Block diagram of PPF Controller

3. Experimental Set-up

The experiments are implemented using existing setup used in [10] and its schematic diagram is shown in figure 3. Three piezoelectric patches are mounted on a 0.5 meter- long aluminum beam. The first piezoelectric is mounted in the middle of the beam and it acts as exciter that excite the beam at its first resonant frequency. The first mode of natural frequency is determined experimentally to be 11.2 Hz. The other two piezoelectric patches are mounted in the collocated fashion near the fix end of the beam and act as sensor and actuator. The sensor provides the feedback signal to the controller whereas the actuator provides the required force to actively control the vibration

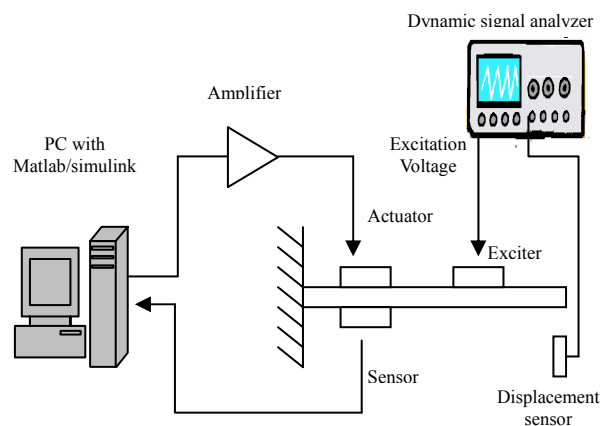


Figure 3: Schematic diagram of experimental setup

The control algorithms are implemented on Pentium 4, PC with 1.5 GHz processing speed under the Matlab/Simulink environment. FL algorithm is first written in C-file, which is then converted into S-function block in Matlab/Simulink. Real-Time-Windows-Target (RTWT) toolbox in Matlab is utilized, in order to enable the connection between the simulink model with the physical system and then execute them in real time. RTWT is a powerful toolbox that enables PC solution for the prototyping and testing of real-time system. It allows the inclusion of I/O blocks into simulink model. From the model, an executable code is generated using the Real Time Workshop (RTW) and C/C++ compiler. With the executable code, simulink can run in external mode to simulate the system in real time.

4. Results and Discussion

Series of experiments are conducted in order to evaluate the effectiveness of FLC. The first series of experiments involve only the beam and the second series involve the use of beam with mass added into it. For each series, FLC is implemented using 9 fuzzy rules and 49 fuzzy rules. Comparative evaluations are made between FL with 9 and 49 rules and between the FLC and PPF controller. The first comparison is aimed to study the effect of number of rules in the performance of the controller and the second comparison is to investigate the effectiveness of FLC in reference to the widely used algorithm.

4.1. Beam system

Figure 4 and 5 show the experimental results of FLC with 49 and 9 rules respectively recorded with the time span of three seconds. It can be observed that vibration is successfully suppressed in both the cases. There is no significant difference can be observed between 9 and 49 fuzzy rules indicating that the FLC with only 9 fuzzy rules is capable of controlling the vibration. Figure 6 shows the result of PPF controller. From the graph, it is obvious that PPF controller performs slightly better than FLC in terms of percentage of amplitude reduction before and after suppression. The transient responses for all cases are almost the same.

4.2 Beam with addition mass added.

Figure 7 and 8 show the results of FLC when the mass is added at the tip of the beam. The maximum amplitude of the signals is lower than the previous cases and this is expected since the addition mass cause the natural frequency of the beam to change. The results indicate that FLC is capable of controlling the vibration even in the presence of uncertainty in the model parameter. Again, there is no significant difference between both graphs in terms of percentage reduction of amplitude. Figure 9 show the result of PPF controller and again it performs better than FL controller in both the cases.

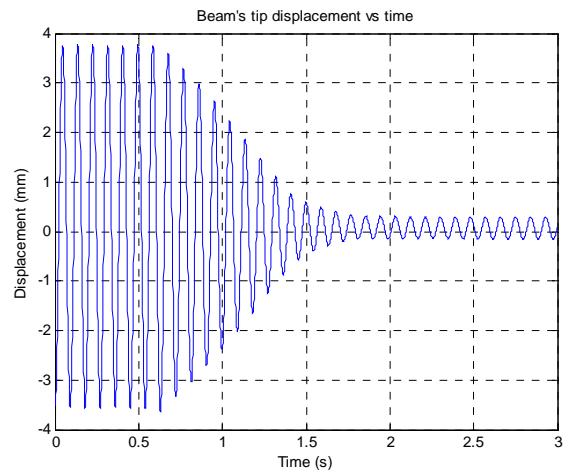


Figure 4: Vibration control using FLC with 49 fuzzy rules

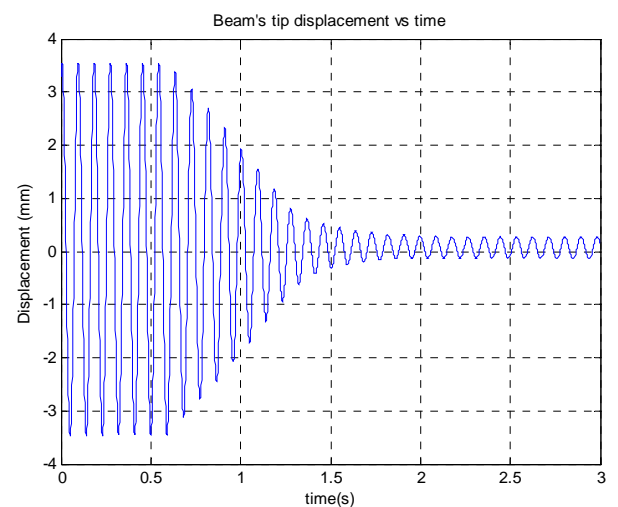


Figure 5: Vibration control using FLC with 9 fuzzy rules

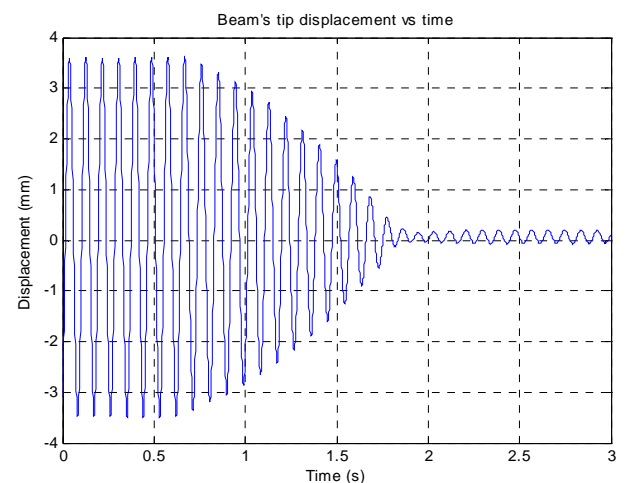


Figure 6: Vibration control using PPF controller

4.3 Comparative assessment

Table 1 summarizes the performance of FLC and PPF in all the cases. From the summary, the following overall observations and conclusion can be made

1. Direct FLC can effectively suppress the vibration of the beam.
2. The number of rules in FLC does not significantly affect the performance of the controller.
3. PPF controller shows superior vibration suppression performance compare to FLC. However the difference is small (less than 8%)
4. Both FLC and PPF are capable of controlling the vibration even in the presence of disturbance/uncertainty in the model parameter

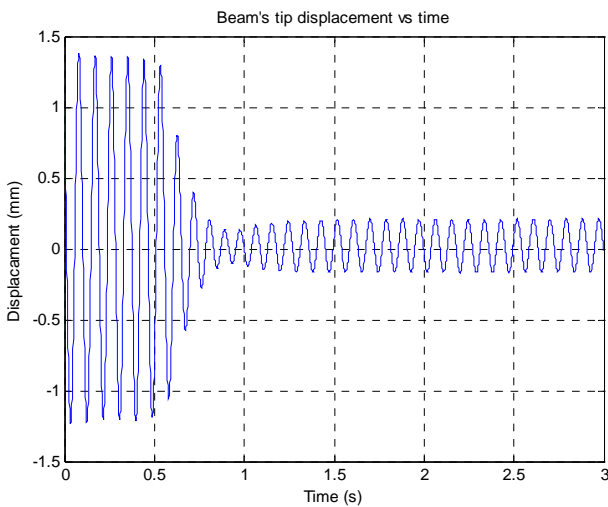


Figure 7: Vibration control using FLC with 49 fuzzy rules (with mass added)

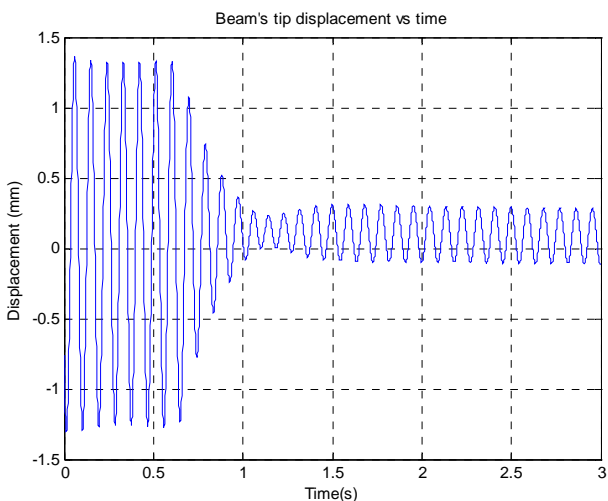


Figure 8: Vibration control using FLC with 9 fuzzy rules (with mass added)

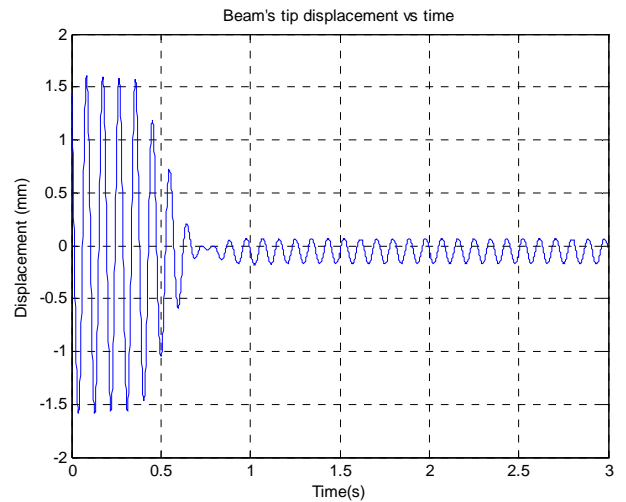


Figure 9: Vibration control using PPF (with mass added)

Table 1. Results of different controllers for beam system

Controller	Maximum amplitude (mm)	Minimum amplitude (mm)	Percentage reduction in amplitude (%)
FLC with 9 rules	3.78	0.23	94
FLC with 49 rules	3.55	0.22	94
PPF controller	3.63	0.14	96

Table 2. Results of different controllers for beam system with added mass.

Controller	Maximum amplitude (mm)	Minimum amplitude (mm)	Percentage reduction in amplitude (%)
FLC with 9 rules	1.37	0.21	85
FLC with 49 rules	1.38	0.20	86
PPF controller	1.61	0.12	93

5. Conclusion

This paper has demonstrated the effectiveness of FLC in suppressing the vibration of cantilevered beam even at the presence of uncertainty in model parameter. FLC is easy to implement and does not require exact mathematical model. The number of rules in the rule-based of fuzzy system has little or no effect on the performance of vibration suppression. Comparative evaluation of FLC showed

inferior performance relative to PPF controller. However, with proper tuning of fuzzy system in terms of its rules and scaling gain, it is believed that FLC will perform better than PPF. Another improvement that can be made on the FLC is to incorporate the learning or adaptive capability.

References

- [1] Tsoukkas, A and VanLandingam, H. 1991. Application of fuzzy logic control to active vibration damping. In *Proceedings of the Eighth VPI&SU*. 371-379 Symposium On Dynamics and Control of Large Structure
- [2] Kwak, M.K and Sciulli, D. 1996. Fuzzy-Logic Based Vibration Suppression Control Experiments On Active Structures. *Journal of Sound and Vibration* 191(1):15-28.
- [3] Sharma, M., Singh, S.P. and Sachdeva, B.L. 2003. Theoretical and Experimental Investigation Of Fuzzy Logic Based Active Vibration Control of Beam. In *Proceedings of DETC'03*. 2109-2117. Chicago, Illinois USA. ASME 2003 Design Engineering Technical Conferences and Computer and Information in Engineering Conference
- [4] Mayhan, P. and Washington, G. 1998. Fuzzy Model Reference Learning Control: A new Control Paradigm For Smart Structures. *Smart Materials and Structures*. 1998: 874-884.
- [5] Gu, H. and Song, G. 2005. Active vibration suppression of a composite I-Beam Using Fuzzy Positive Position Control. *Smart Materials and Structures*. 2005: 540-547
- [6] Zeinoun, I.J. and Khorrarni, F. An adaptive control scheme Based on Fuzzy Logic and Its Application To Smart Structures. *Smart Materials and Structures*. 1994: 266-276
- [7] Passino, K.M and Yurkovich, S. 1998. *Fuzzy Control*. Reading, Mass.: Addison-Wesley
- [8] Goh, C. and Caughey, T.K. 1985. On the stability problem caused by infinite actuator dynamics in collocated control of large space structure. *International Journal of Control*. 41:787-802
- [9] Lingfeng, W. 2003. Positive Position Feedback Based Vibration Attenuation For A flexible Aerospace Structure Using Multiple Piezoelectric Actuators. In *IEEE conference proceeding DASC'03*. 7.C.1-1 – 7.C.1-10. The 22nd Digital Avionics System Conference.
- [10] Asan Gani, A.M. 2003. Active Vibration Control Using Piezoelectric Materials. Msc Dissertation, Department of Mechatronic Engineering, International Islamic University.

Mode-I Interlaminar Fracture Toughness of Hand Lay-Up Plain-Weave Woven GFRP/Unsaturated Polyester Laminate

Mohd Aidy Faizal^{1*}, Yeo Kiam Beng¹, Mohd Noh Dalimin¹

¹ Center of Materials and Minerals (CMM)

Universiti Malaysia Sabah, Beg Berkunci No. 2073, 88999 Kota Kinabalu, Sabah, Malaysia

Tel. 088-320000 ext: 3543 E-mail: aidyfaizal_7513@yahoo.com

Abstract

This paper describes an experimental study on the interlaminar mode-I fracture behavior of hand lay-up plain-weave woven GFRP laminate by using a fracture energy method of analysis. The mode-I interlaminar tests were conducted using the double cantilever beam (DCB) geometry tested at room temperature (23°C/297.15K) with a 3mm/min constant loading rate. The tests were performed with reference to the ASTM D5528- Standard test method for mode-I interlaminar fracture toughness of fiber-reinforced polymer matrix composite. The fracture energy analysis for the DCB test specimen enables the energy release rate or fracture energy, G_{Ic} , to be evaluated by linear elastic fracture mechanic (LEFM). From the load-displacement trace, the data obtained is analyzed using several LEFM methods including corrected beam theory and the compliance method. Both analyses produce the fracture toughness energy curve, which indicates the rising R-curve effect. The increasing R-curve is significantly related to the deformation of fiber and fiber bundle bridging, breakage and tow separation phenomena, which arises between the cracked planes.

Keywords

Woven GFRP Laminate , Mode-I, Double Cantilever Beam (DCB), Fracture Toughness

1. Introduction

Composite materials are ingenious invention that provided immense benefits in the application of engineering design and structure. Nowadays, these materials are used in a very wide range of industrial applications. Laminated fiber-reinforced composite materials are commonly used in the marine industry, ducting and piping industries and many others due to their good environmental resistance, better damage tolerance for impact loading, and at the same time having a high specific strength and stiffness [1]. Fiber woven fabrics are probably the most commonly used textiles in structural applications. It can also be offered as a low cost composite manufacturing.

Mechanical properties can be considered the most important of all the physical and chemical properties for most application [2]. The mechanical properties of woven fabric composites, such as strength and stiffness, are strongly determined by the weave parameters (weave geometry, yarn size, yarn spacing and yarn crimp), the

laminate parameters (fiber orientations and overall fiber volume fraction), and the inherent material properties of fiber and matrix. However, delamination of laminated composite materials is a critical problem, whereby interlaminar failure during routine services becomes a major concern in structural composite design [3]. Resistance to delamination or fracture toughness is one of the most important mechanical properties of fiber composite. The delamination may grow giving rise to a progressive stiffness reduction, material degradation, and even premature fracture of the composite component.

Studies of elastic failure mechanisms provide us a way to be critical on the development of material industries. The mechanics of fracture by Linear Elastic Fracture Mechanic (LEFM) on composite material has been well developed in recent years and proven to be an efficient tool for evaluation of interlaminar failure. The analysis based on the critical fracture toughness energy value G_{Ic} is commonly adopted for the calculation of interlaminar fracture toughness behavior. The weakest mode of the composite failure has been known to be the Mode-I interlaminar failure. Commonly employed standard test geometry, BS ISO 15024 and ASTM D5528., can be used for characterization of Mode-I interlaminar toughness.

* Corresponding Author. E-mail: aidyfaizal_7513@yahoo.com, Tel: +60-88-320000 (3543)

Fracture toughness can be determined numerically in many ways, affected by many parameters and the test can be conducted by using a wide range of specimen types. The test procedures are currently well established for unidirectional laminates in ASTM D5528 [4] and BS/ISO 15024 [5]. However, practical applications usually involve multidirectional laminates. Therefore, it is essential to study the interlaminar fracture toughness of multidirectional and woven composites [6]. The double cantilever beam (DCB) geometry has been widely used in this study for the characterization of delamination resistance or fracture energy of laminates.

The fracture toughness energy value, G_{Ic} of woven composite has been commonly reported to be higher than those of unidirectional laminates due to complex damage morphology. Other workers [7] reported that interlaminar toughness of reinforced composites is affected by variation in the matrix toughness through increasing the matrix thickness between adjacent fiber space, such as decreasing the fiber content. Investigation on crack propagation [8] in DCB test of multidirectional laminates found major difficulty with multidirectional specimens possessing high tendency to intralaminar cracking and delamination jumping between neighbor interfaces. This work is also similarly supported by other workers [2], which showed that the woven composite may yield significantly greater scatter and unique R -curves associated with varying toughness within and away from the interlaminar resin pockets as the delamination grows.

Focus on present work has been undertake to investigate the interlaminar fracture toughness energy value, G_{Ic} of woven GFRP/Unsaturated Polyester composite material using the fracture mechanics approach, including modified beam theory (MBT) and compliance analysis.

2. EXPERIMENTAL METHOD

The experimental method describes in detail the materials and its fundamental constituents, the specimen preparation for the fabrication of GFRP composite panels, and the experimental test method that has been carried out.

2.1 Materials

Plain-weave type woven glass fiber, commercial code: NISER (EWR-600B) has been used as the GFRP material for the purpose of this investigation. The basis of EWR-600B is the glass fabric of E-glass, according to the 0.5% content of $\text{Na}_2+\text{K}_2\text{O}$. The E-glass plain weave has been produced by interlacing warp thread and fill thread as illustrated in the Figure 1.

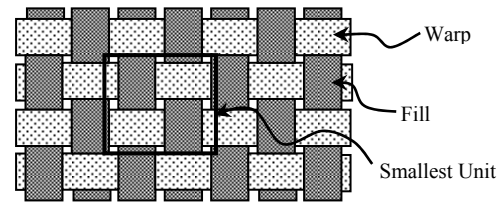


Figure 1. Typical plain-weave woven E-glass.

The EWR-600B fiber has a round cross sectional shape approximately $9\mu\text{m}$ and chemically contains less than 1% alkali. It has a specific mass of $600\text{g}\cdot\text{m}^{-2}$ and dry thickness of 0.47 mm. The EWR-600 has a round cross sectional shape approximately $9\mu\text{m}$ and chemically contains less than 1% alkali. The basic physical properties of EWR-600B was describe in Table 1.

Table 1. Physical properties of EWR-600B

Description		Value
Code		EWR 600B
Width (mm)		1000
Linear density (tex)	Warp	1750
	Weft	1750
Fabric density (end/cm)	Warp	1.97
	Weft	1.57
Area weight (g/m^2)		620
Moisture content (%)		< 0.15
Warp size (mm)		2.48
Weft size (mm)		2.10
Yarn spacing (mm)		1.08
Dry thickness (mm)		0.467

The matrix used was unsaturated polyester resin, commercially coded 268BQTN and supplied by SHCP. The resin has excellent mechanical strength, good rigidity, light weight and outstanding durability. The polyester 268BQTN used having a density of $1.21\text{g}/\text{cm}^3$. This resin was cured by the application of catalyst hardener, Butanox M-50 (methyl ethyl ketone peroxide, solution in dymethyl phthalate). Cure of the polyester resin was achieved using 0.74% by mass of catalyst. The resin was mixed with the hardener in the ratio of 400: 3 by mass.

2.2 Lay-Up and Curing Process

An open rectangular mold made out of 4mm thick mild steel plates is used for the lay-up and curing of woven GFRP. The rectangular mold has a dimension of 435mm x 305mm x 45mm.

Laminate preparation is through a hand lay-up process at room temperature. A plain weave woven ply with fibers at 0° and 90° direction is layered with symmetrically arrangement at the mid-plane of the laminate. Each layer of the woven fiber fabric is layered

over each other with a layer of resin between. The completed lay-up is then subjected to a uniform pressure of 70.1kg/m^2 through a piece of mild steel-plate. The whole set up is then left for curing over two days, Figure 2 shows the lay-up processes.

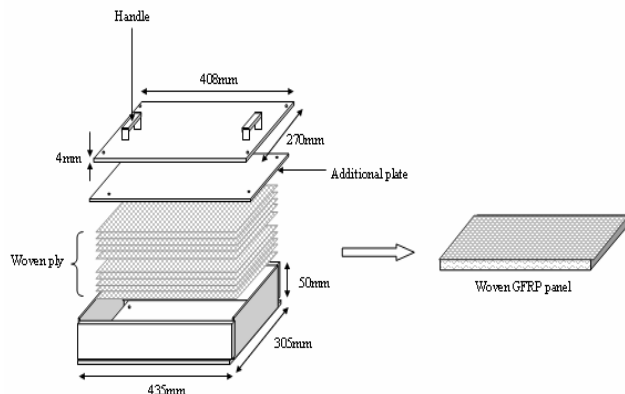


Figure 2. Fabrication of woven GFRP panel

2.3 DCB Sample

The DCB or double cantilever beam specimen is fabricated with a starter crack in using an aluminum film of $13\mu\text{m}$ thickness placed at the mid-plane of the laminate during fabrication process. All DCB samples are produced from the composite panels by cutting into rectangular strips using a diamond circular saw. The mechanical damages caused by the cutting are removed from the straight edges by polishing. Load mounting on the specimen is achieved through the use of rectangular mild steel end-blocks, which are bonded to one end of the specimen using high strength Devcon white epoxy adhesive. The DCB specimens are produced according to the ASTM D5528-Standard test method for mode-I interlaminar fracture toughness of fiber-reinforced polymer matrix composite [2]. The geometry of the DCB specimen consists of a rectangular uniform thickness as illustrated in the Figure 3.

Dimensions of the specimen have been selected according to the standard test method [4] and [5] specifications, and as well as other publications [3-12] are referred. Perin and workers [9] reported that improper selection of thickness of the DCB specimens will required the need for large deflection correction, which is the cause and result of large displacement and end block rotation. The references [2] and [3] provided the guidelines for choosing a specimen thickness that will yield negligible displacement corrections based on the anticipated interlaminar fracture toughness.

The geometric parameters of the DCB specimen are illustrated in the Figure 3. The b is the specimen width, h the specimen thickness, a_0 the initial delamination

length, a the total delamination length, A the insert length, l the specimen length, l_1 the distance from center of loading pin to mid-plane of specimen, l_2 the distance from center of loading pin to edge of end block, l_3 the end block length, and H the end block thickness.

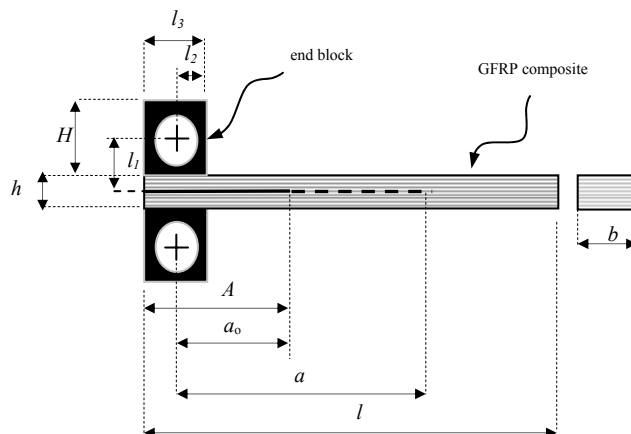


Figure 3. Specimen geometry and parameters

3. STRAIN ENERGY RELEASE RATE

The linear elastic fracture mechanics (LEFM) analysis for analyzing the data contained in the load-displacement traces can be tackled by different approaches. The various methods available for calculating mode-I strain energy release rate from DCB geometries could result in different values [10]. In this study however, a load/displacement plot provided data for each crack length used for the calculation of G_{Ic} obtained from the experimental beam theory and compliance method (Berry's method).

The experimental beam theory expression for G_{Ic} perfectly built-in (that is, clamped at the delamination front) of DCB specimen can be modeled as

$$G_{Ic} = \frac{3P\delta}{2ba} \quad (1)$$

where P is the load, δ the load point displacement, b the specimen width, and a the delamination length. However, this equation could over estimate the G_{Ic} by imperfect built-in that can be due to the rotation at the delamination front. Correction of this rotation is to treat the DCB as if it contains a slightly longer delamination, and this led to the introduction of Modified Beam Theory (MBT) analysis developed as

$$G_{Ic} = \frac{3P\delta}{2b(a + |\Delta|)} \quad (2)$$

where Δ is determined experimentally by generating a least square plot of the cube root of compliance, $(C/N)^{1/3}$ as a function of delamination length. The parameter- C is the compliance or also the ratio of the load point displacement to the applied load, δ/P , and N is the loading block correction. The determination for this correction is illustrated in Figure 4.

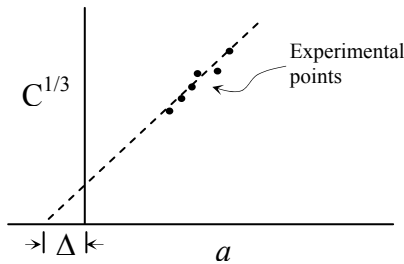


Figure 4. Determination of the Δ -correction

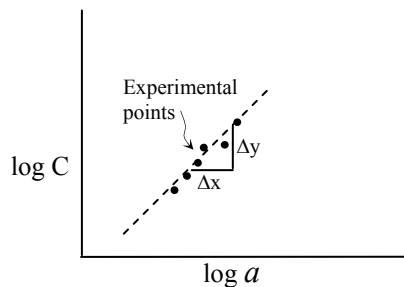


Figure 5. Determination of compliance calibration

Compliance Method is obtained by generating a least square plot of $\log \delta_i/P_i$ versus $\log a_i$ in order to visually observe an onset value, such that the propagation values of fracture energy can be derived as

$$G_{IC} = \frac{n P \delta}{2 b a} \tag{3}$$

where n is the Δ_y/Δ_x , and the Δ_y and Δ_x are defined from the slope of the line $\log \delta_i/P_i$ versus $\log a_i$ graph as shown in Figure 5. As a result, the critical energy release rate G_{IC} is determined by multiplying the F/N factor to the G_{IC} in equation (1-3), such that

$$F = 1 \frac{3}{10} \frac{\delta}{a}^2 \frac{2}{3} \frac{\delta l_1}{a^2} \tag{4}$$

$$N = 1 \frac{l_2^3}{2} \frac{9}{8} 1 \frac{l_2^2}{a} \frac{\delta l_1}{a^2} \frac{9}{35} \frac{\delta}{a}^2 \tag{5}$$

where F is the large displacement effect and N is the loading block correction.

4. RESULT AND DISCUSSION

The typical load-displacement characteristic curve for the interlaminar fracture of woven GFRP is as shown in the Figure 6. The delamination crack growth process is a stable propagation with decreasing load for increasing displacement. Slow increment of initiation of crack in the first few millimeters of crack propagation has been observed. The initiation develops over a short 3mm distant of crack, which is displayed along with a slight non-linear load-displacement characteristic. This is also attributed to the larger displacement caused by initial rotation of the arm-beam. Subsequently, the typical crack propagation process is reflected on the load-displacement trace as a decreasing load shift with increasing displacement.

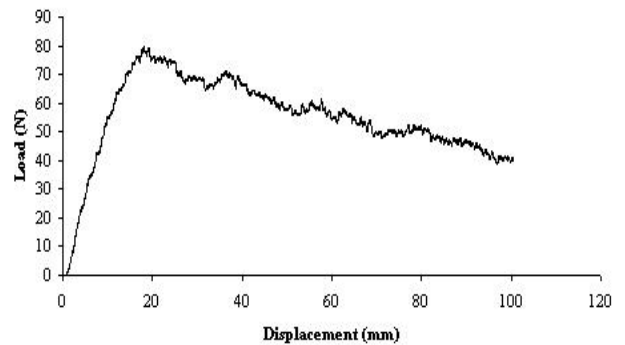


Figure 6. Load versus displacement

Fracture energy toughness evaluation and characterization using the MBT and the Compliance analysis are computed for comparison in Figure 7. The delamination-resistance characteristics typical of the Modified Beam Theory and Compliance analysis indicate an increasing R-curve. Stable crack propagation through initiation is followed by unstable crack growth of the R-curve. The visually observed initiation fracture energy value is labeled VIS, while the propagation value is indicated as PROP in Figure 7. The coefficient of variation for the characteristic R-curve by the MBT is 4%, and the Compliance method has about 4.7%. However, the shift of the two set of R-curve by each analysis is explained by the effect of cracked arm rotation, which has not been corrected in the Compliance method of analysis.

The increasing R-curve is significantly related to the deformation of fiber bridging and tow separation phenomena. The fibers and fiber bundles between cracked planes are found to develop bridges, breakages

and separation leads to the instability of crack propagation recognized as the rising R-curve.

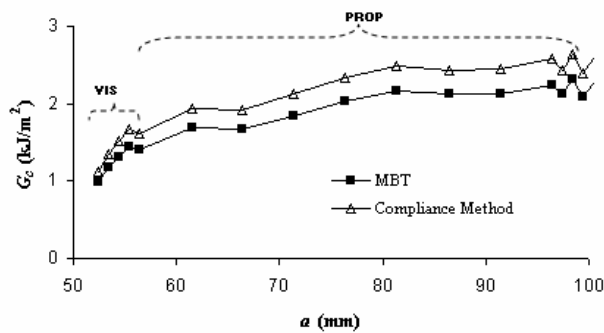


Figure 7. Typical delamination-resistance curve (R-curve) showing increasing delamination resistance with delamination length

CONCLUSIONS

Studies on the mode-I interlaminar fracture toughness of the woven GFRP composite was conducted. The MBT and Compliance data reduction methods produced a shift in their fracture toughness energy values or the R-curves by about 8%. The correction multiplier F/N has shown to be more effective for near initiation or stable crack initiation phase. However, in the propagation phase the multiplier does not effectively corrected the variation between the two model approaches. This may be explained by the ineffectiveness of the multiplier term to correct for non-linearity characteristic in stiffness flexibility of beam-arm and the material properties of woven GFRP composite laminate.

Acknowledgements

The authors wish to express their sincere appreciation to the *Centre of Materials and Minerals, Universiti Malaysia Sabah*, for their support in this project.

References

- [1] Sutherland, L.S, Soares, C.G. Effect of laminate thickness and of matrix resin on the impact of low fiber-volume, woven roving e-glass composites. *Compos Sci Technol* 2004;64:1691-1700.
- [2] Nielsen, L.E, Landel, R.F, Faulkner, L.L, editor. *Mechanical properties of polymers and composites*. 2nd ed. New York: Marcel Dekker Inc.; 1994.
- [3] Yeo, K.B, Mohd Noh Dalimin. Mode-I interlaminar fracture toughness of unidirectional carbon fiber epoxy composite. *Borneo Science* 2001;10:1-9.
- [4] ASTM Standard D5528-01. Standard test method for mode-I interlaminar fracture toughness of fiber-reinforced polymer matrix composite. American Society for Testing Materials, West Conshohocken, PA, 2001.
- [5] Fiber-reinforced plastic composites-Determination of mode-I interlaminar fracture toughness, GIC, for unidirectionally reinforced materials. (BS ISO 15024). United Kingdom British: Standard. 2001.
- [6] de Morais, A.B, de Moura, M.F, Marques, A.T, de Castro, P.T. Mode-I interlaminar fracture of carbon/epoxy cross-ply composite. *Compos Sci and Technol* 2002;62:679-686.
- [7] Srivastava, V.K, Hogg, P.J. Damage performance of particles filled quasi-isotropic glass-fiber reinforced polyester resin composites. *J Mater Sci* 1998;33:1119-1128.
- [8] de Morais AB, de Moura MF, Goncalves JPM, Camanho PP. Analysis of crack propagation in double cantilever beam test of multidirectional laminates. *Mechanics of Materials* 2003;35:641-652.
- [9] Perrin F, Bureau MN, Denault J, Dickson, JI. Mode-I interlaminar crack propagation in continuous glass fiber/polypropylene composites: temperature and molding condition dependence. *Compos Sci and Technol* 2003;63:597-607.
- [10] Ashcroft IA, Hughes DJ, Shaw SJ. Mode-I fracture of epoxy bonded composite joints: 1. Quasi-static loading. *Int J Adhesion & Adhesives* 2001; 21:87-99.
- [11] Shivakumar K, Chen H, Abali F, Davis C. A total fatigue model for mode I delaminated composite laminates. *Int J Fatigue* 2006; 28:33-42.

Fabrication of Polymer Light Emitting Diodes with ITO/PVK:TPP/Alq₃/Al Structure

C.C. Yap^{1*}, M. Yahaya¹ and M.M. Salleh²

¹Faculty of Science and Technology

²Institute of Microengineering and Nanoelectronics (IMEN)

Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

Tel:+60-3-89213560, Fax:+60-3-89213777, E-mail: chichin83@yahoo.co.uk

Abstract

Polymer light emitting diodes with the configuration of ITO/PVK:TPP/Alq₃/Al were fabricated where indium tin oxide (ITO) was used as anode, poly(9-vinylcarbazole) (PVK) as host, tetraphenylporphyrin (TPP) as red dopant, tris(8-hydroxyquinoline) aluminium (Alq₃) as electron-transporting green emitter layer and aluminium (Al) as cathode. The performances of the devices with various TPP doping concentrations by weight in PVK host were examined through the current-voltage (I-V) curve and electroluminescence (EL) spectrum. The EL spectrums as well as the operating voltage were found to be dependent on the TPP doping concentration.

Keywords:

Polymer light emitting diodes, electroluminescence, PVK, TPP, Alq₃.

1. Introduction

Polymer light emitting diodes (PLEDs) have attracted considerable interest due to their easy processability, flexibility, low cost, low operating voltages, wide viewing angles, tunability of the color emission, fast response time, and ease of forming large area [1]. The dopant/host emitter system is a common strategy to tune the emission color of PLEDs. The energy transfer and charge trapping from the host to the dopant molecule allows the emission to originate from the dopant and the emission properties are dominated by the dopant [2]. Moreover, the dopant/host system is found to enhance the EL efficiency of PLEDs [3].

Many dyes have been utilized as emitting center in red PLEDs. Among them, tetraphenylporphine (TPP) has been doped into poly(9,9-dioctylfluorene) to produce red PLEDs via efficient energy transfer [3]. In this paper, we report an alternative dopant/host emitter system where TPP was doped into poly(9-vinylcarbazole) (PVK).

PVK is one of the most frequently used polymeric hosts, due to its excellent film-forming and hole-transporting properties [4]. Since PVK is mainly hole-transport, owning very limited electron-transport capabilities, it is expected that in the single-layer structure, the carrier recombination zone be closer to the cathode due to the highly unbalanced transport properties of holes and electrons [5]. Therefore, an electron-transporting material, tris(8-hydroxyquinoline) aluminium (Alq₃) was employed

to transport electrons into dopant/host emitting layer, and separate the emitting layer from cathode to prevent luminescence quenching by the cathode surface [6].

We have successfully fabricated polymer light emitting diodes (PLEDs) with ITO/PVK:TPP(0.5-5wt%)/Alq₃/Al structure where PVK was used as polymeric host, TPP as red dopant, Alq₃ as electron transporting green emitter layer. The EL spectrums as well as the operating voltage were found to be dependent on the TPP doping concentration. The light observed was the combination from TPP and Alq₃ emission even at high concentration of TPP. The mechanism of light emission was discussed based on the energy level diagram.

2. Methods

The host polymer PVK, having a high weight-average molecular weight of 1,100,000 g/mole was purchased from Aldrich Co. Inc. Besides, the red dopant, TPP and the electron transporting green emitter, Alq₃ were purchased from Tokyo Kasei Kogyo Co. Ltd. The molecular weight of Alq₃ is 459.43g/mole with the melting point of 300-320°C. All materials were used as received without further purification.

The two-layer PLEDs with ITO/PVK:TPP/Alq₃/Al structure were fabricated as shown in Figure 1. The ITO-coated glass substrates were etched and patterned to serve as anode. The substrates were cleaned with 2-propanol and acetone in an ultrasonic bath, respectively, and then dried with a hair dryer. 1,2-dichloroethane solutions containing 10mg/mL of 0.5-5wt% TPP in PVK were prepared. The solutions were then spin-coated onto the ITO with a

* Corresponding Author. E-mail: chichin83@yahoo.co.uk, Tel:+60-3-89213560, Fax:+60-3-89213777

typical spinning speed and time at 2000rpm for 40s. After drying the polymer layer, the Alq₃ layer was vacuum deposited to a thickness of 3nm. Lastly, 150nm aluminium was then deposited by electron gun evaporation technique without breaking the vacuum.

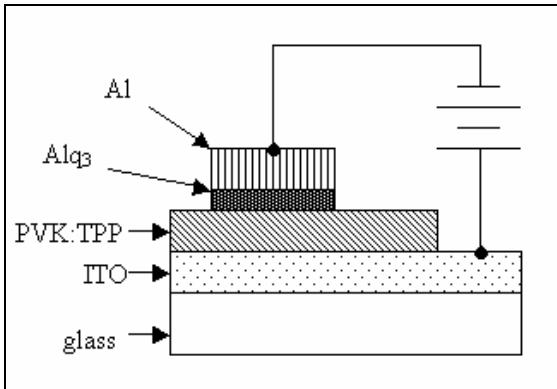


Figure 1. The PLED with ITO/PVK:TPP/Alq₃/Al Structure

The photoluminescence(PL) properties of PVK, Alq₃ and TPP(PMMA as host matrix) were investigated by depositing the thin films on quartz substrates. The PL spectra were investigated with Perkin Elmer LS55 luminescence spectrometer. In addition, Keithley 238 source measurement unit was used to measure the electrical characteristics of the devices, while the electroluminescence (EL) spectra were obtained with Ocean Optic HR2000 spectrometer. All measurements were carried out at ambient atmosphere.

3. Results and Discussion

The PL spectra of the materials used are depicted in Figure 2. As shown in Figure 2, the TPP exhibited a narrow red emission peak at 650nm. Upon photo-excitation the PVK and Alq₃ emitted blue light peaked at 388nm and green light peaked at 525nm respectively.

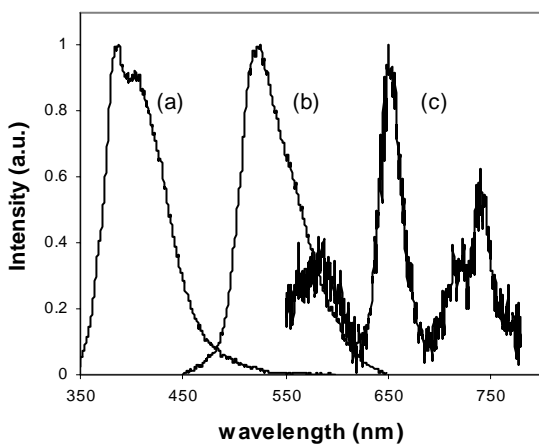


Figure 2. PL of PVK (a) in Thin Film, PL of Alq₃ (b) in Thin Film and PL of TPP (c) in PMMA Host (5wt%)

Figure 3 shows the normalized EL spectra of devices ITO/PVK:TPP/Alq₃/Al with four TPP doping concentrations at the same driving voltage of 28V. In all the spectra, there was a red luminescence band whose peak was 650 nm due to TPP emission. However, the spectra also showed a green luminescence band due to Alq₃ emission. Emission from PVK cannot be observed at any doping concentration.

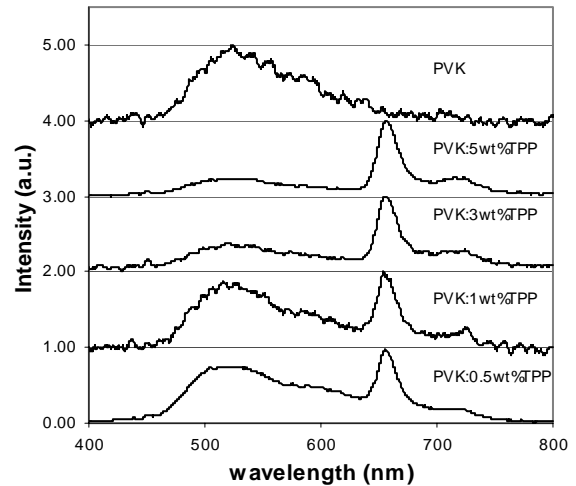


Figure 3. Normalized EL Spectra of Devices at Different TPP Concentrations

Figure 4 shows a proposed energy level scheme for the ITO/PVK:TPP/Alq₃/Al devices. The Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO) energy level for all the materials used were taken from reference [7]. Based on the energy diagram, the jump of electrons from Alq₃ to PVK needs to overcome a potential barrier of 0.8eV while the jumping potential barrier from Alq₃ to TPP is less than zero. Therefore, electrons are much easily transported directly from Alq₃ to TPP than from Alq₃ to PVK. As the TPP doping concentration increased, the electron-hopping rate from Alq₃ to TPP increases too [8]. Enhancement of the ratio of the emission from TPP to that from Alq₃ with the increase of TPP doping concentration supports this point of view.

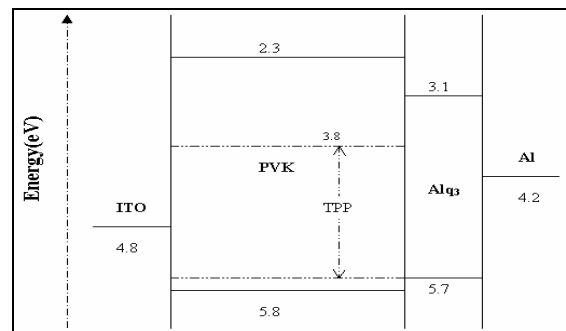


Figure 4. Schematic Energy Level Diagram of the ITO/PVK:TPP/Alq₃/Al Devices

Besides, we also fabricated a PLED consisting of a neat PVK to study the role played by the TPP dopant. Here, a large barrier (0.8 eV) exists between the LUMO of PVK and Alq₃, larger than the barrier between the HOMO of PVK and Alq₃. The electrons will be blocked at the interface and the holes can transport into Alq₃ layer. Thus, a strong Alq₃ EL emission was observed since most of the recombination takes place in the Alq₃ layer where the electrons are confined.

Figure 5 shows the effect of TPP doping concentration on the current-voltage characteristics at room temperature. The turn-on voltage for the ITO/PVK/Alq₃/Al device is the highest. This indicates that injection of electron into LUMO of PVK needs higher energy as compared to direct injection of electron into LUMO of TPP. As the concentration of TPP increases, the current decreases at a given bias voltage. This result indicates that TPP form a hole trap state within PVK due to its higher lying HOMO level as shown in Figure 4.

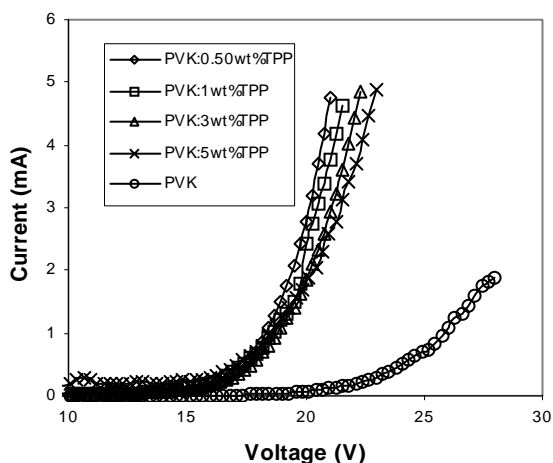


Figure 5. Current-voltage Characteristic of Devices at Different TPP Concentrations

4. Conclusion

The polymer light emitting diodes with the ITO/PVK:TPP/Alq₃(3 nm)/Al structure has been successfully fabricated. The EL spectra and the operating voltage of the devices were found to be dependent on the TPP doping concentration. However, the turn-on voltage of the device was too high (15V). After optimizing the device structure such as using lower work function cathode, introducing thin buffer layer and adjusting the polymer layer thickness, further reduce of the turn-on voltage is expected.

Acknowledgements

This work has been carried out with the support of Malaysian Ministry of Science, Technology and Innovation, under the IRPA grant 03-02-02-0067-SR0007/04-04.

References

- [1] Gupta, D.; Katiyar, M.; and Deepak. 2005. Various Approaches to White Organic Light Emitting Diodes and Their Recent Advancements. *Optical Material* 28(4): 295-301.
- [2] Uchida, M.; Adachi, C.; Koyama, T.; and Taniguchi, Y. 1999. Charge Carrier Trapping Effect by Luminescent Dopant Molecules in Single-layer Organic Light Emitting Diodes. *Journal of Applied Physics* 86: 1680-1687.
- [3] Virgili, T.; Lidzey, D.G.; and Bradley, D.D.C. 2000. Red-light-emitting Diodes via Efficient Energy Transfer from Poly(9,9-dioctylfluorene) to Tetraphenylporphyrin. *Synthetic Metals* 111: 203-206.
- [4] Kido, J.; Hongawa, K.; Okuyama, K.; and Nagai, K. 1994. White light-emitting Organic Electroluminescent Devices Using the Poly(N-vinylcarbazole) Emitter Layer Doped with Three Fluorescent Dyes. *Applied Physics Letters* 64: 815-817.
- [5] Kido, J.; Hongawa, K.; Okuyama, K.; and Nagai, K. 1993. Bright Blue Electroluminescence from Poly(N-vinylcarbazole). *Applied Physics Letters* 63: 2627-2629.
- [6] Kim, K.; Lee, D.W.; and Jin, J. 2000. Electroluminescence Properties of Poly[2-(2'-ethylhexyloxy)-5-methoxy-1,4-phenylenevinylene] /Tris 8-hydroxyquinoline aluminum Two-layer Devices. *Synthetic Metals* 114: 49-56.
- [7] Ohmori, Y.; Kajii, H.; Sawatani, T.; Ueta, H.; and Yoshino, K. 2001. Enhancement of Electroluminescence Utilizing Confined Energy Transfer for Red Light Emission. *Thin Solid Films* 393: 407-411.
- [8] Chen, X.; Liu, M.; Xu, Z.; Hou, Y.; Teng, F.; and Xu, X. 2004. Carrier Transport Assisted by Dopants in Doped Poly(N-vinylcarbazole) Light-emitting Diodes. *Journal of Physics D: Applied Physics* 37: 1007-1011.

Identifying Meshing Characteristics To Model Crack Tip Region For Linear Fracture Mechanics Application

E.H. Lim^{*}, K. B. Yeo, Mohamed Harimi

Centre of Materials and Minerals

Universiti Malaysia Sabah, Locked Bag 2073, 88999 Kota Kinabalu, Sabah, Malaysia

Tel: +6-088-320000, Ext: 3543, E-mail: ehlim@pasca.ums.com.my or stan_hel@hotmail.com

Abstract

The importance of meshing technique in numerical fracture analysis is emphasized in this paper. Investigation on cracked models with and without localized region is carried out to show the significance of the localized stress analysis. The model is a standard compact tension, a common geometry used in standard experimental testing to obtain material fracture toughness. Toughness value obtained via ASTM Standard coded E-399 is referred for the investigation of numerical toughness, represented by crack-tip stress intensity factor, which is used to correlate to the analyzed meshing techniques. The localized meshing is found to yield convergence and stability better than the large scale meshing. The computation has also found that the localized region enclosed with five equally distributed rings produces the lowest 4.5% variation from the experimental value.

Keywords:

Meshing, localized region, fracture toughness, crack-tip stress intensity factor.

1. Introduction

Material technology has been the main foundation of human civilization and development [1]. The advances of navigations, aviations, industrialization and so forth have witnessed how mankind can make use of materials in their daily applications. The increasing knowledge of material behaviors has further the engineering for more superior utilization. Inevitably, materials are subjected to a life-span and service operation that can fail and even fracture within the elastic limit [2]. This shortcoming has caused losses of innumerable human lives and properties through the centuries. Such events like the sinking of more than a thousand Liberty Ships, failure of comet airliner, numerous breakdowns of bridges, railroad axles and boilers have awoken the seriousness towards the deficiency in understanding of material behavior [3].

Committees and laboratories have been developed in searching solutions to prevail over material failure, while at the same time allowing material technology to keep on progressing. Throughout decades of strenuous efforts from researchers, finally a material discipline that can describe more thoroughly of material failure is developed. In 1948, Irwin published his works on the principle used to analyze material fracture, particularly in crack phenomena, namely the famous fracture mechanics. This discipline has since then been employed in fracture study for most of the existing materials until today [4].

Conventional analytical method appeals to physics and engineering principles to derive differential equations relating the variables of interest in a system [5]. However this method has limitation in solving excess or large number of variables. As the testing is costly and time-consuming, numerical fracture analysis by Finite Element Method has emerged as a very useful approach for fracture investigation. The pioneer development of numerical fracture analysis dates back to 1960 when R.W. Clough first applied the modeling tool in plane stress analysis [6]. One of the major features in finite element analysis that controls the modeling solution is the mesh generation. It defines the reaction of numerical model to the loading conditions, material properties, and constraints, which best simulates the real problem [7]. An accurate and a good crack tip modeling by numerical approach also requires a combination of optimization of meshing parameters such as element size, shape, number, and region of interest.

2. Experimental Test

The aim of the experiment is to obtain the plane strain fracture toughness of a medium carbon-steel material. The material is designed and machined in accordance to ASTM coded E-399 for straight-notch standard compact tension model as illustrated in the Figure 1.

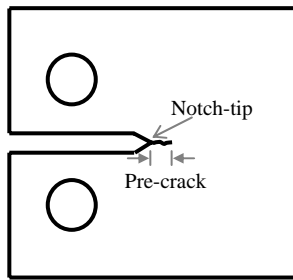


Figure 1. Geometry for Standard Compact Tension Model.

The dimensions of specimen is designed to fulfill the basic condition of plane-strain relating the crack length a , specimen thickness B , and yield stress σ_y with the mode-I plane strain fracture toughness K_{IC} .

$$a, B \geq 2.5(K_{IC} / \sigma_y)^2 \quad (1)$$

For K_{IC} testing, the notch-tip of the specimen is pre-cracked by fatigue loading to yield a sharp crack as close as possible in order to generate the elastic fracture with limited or no local plasticity. For valid testing [8], a set of loading conditions must be achieved as in the following.

For pre-crack test:

$$-1.0 \leq K_{\min} / K_{\max} \leq +0.1 \quad (2)$$

$$(K_{\max})_{\text{initial stage}} \leq 80\% K_{IC} \quad (3)$$

$$(K_{\max})_{\text{terminal stage}} \leq 60\% K_{IC} \quad (4)$$

For K_{IC} test:

$$0.55 \leq dK/dt \leq 2.75 \text{ MPa.m}^{0.5}/\text{s} \quad (5)$$

Upon validating the tests, the toughness value is calculated from the following equation, which yields an average value of $32.586 \text{ MPa.m}^{1/2}$ at no more than 3% coefficient of variation.

$$K_Q = \frac{P_Q}{BW^{0.5}} \cdot f(a/W) \quad (6)$$

where P_Q is obtained from the load-displacement graph by offsetting a 5% secant linear slope, whereas $f(a/W)$ is a geometrical factor calculated from the dimensionless equation below.

$$f(a/W) = \frac{(2 + a/W) \left(\begin{array}{l} 0.886 + 4.64a/W - 13.32a^2/W^2 \\ + 14.72a^3/W^3 - 5.6a^4/W^4 \end{array} \right)}{(1 - a/W)^{3/2}} \quad (7)$$

B and W are the specimen thickness and width, with pre-stated values of 26mm and 52mm respectively.

3. Finite Element Meshing

This paper is confined to the investigation of meshing technique that emphasizes the necessity of localized stress analysis. A single element-type of two-dimensional quadrilateral elements, ANSYS 9.0 coded PLANE82, is used for both meshing with and without localized-region. Since the model geometry exhibits the symmetrical characteristic, a half numerical model is created to represent the 2-dimensional plane. The key parameter in this study is the numerical toughness, represented by the crack-tip stress intensity factor K_{ct} . Accordingly relatively smaller elements are concentrated in this region of interest that initiated from the crack-tip [9]. This arrangement constitutes the 'spider web' configuration, supported by Anderson [10] as the most efficient mesh type for crack-tip stress analysis. It has the capability to generate refined meshes about the vicinity of the crack-tip and construct meshes from fine to coarser degree as illustrated in the Figure 2.

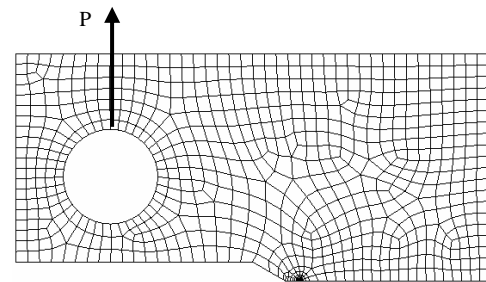


Figure 2. Numerical Model Adopted.

3.1. Meshing-without-localized-region (M-WOLR)

The half-model of the Compact Tension geometry in the Figure 2 is employed for a global computational analysis. The model is simply meshed by setting the concentration point at the crack-tip. Several degrees of meshing refinement are carried out on the variation of sizes of first row elements, R , in order to ensure the accuracy and precision of the numerical K_{ct} . The meshing technique is investigated over the row R from 100 to 500 microns.

3.2. Meshing-with-localized-region (M-WLR)

The model geometry of Figure 2 is now adjoined with a small localized-region around the crack-tip. The size of region is designed in the form of semi-circle set as 130 microns, 0.5% of crack length a , and small enough for the local stress analysis. This localization characteristic enables the generation of supplementary organized fine meshes about the vicinity of the tip that structures better arrangement of nodes. The analyzed R also scaled from 100 to 500 microns as a means of comparison between the two techniques.

In the case of localized technique, the distribution of rings within the localized region has significant effect on the numerical K_{ct} . Region enclosed with more rings

provides higher degree of node density around the vicinity of the crack-tip. This characteristic is demonstrated in the Figure 3 that illustrates the distribution of nodes by different levels of density within the localized region. However the variation of K_{CT} against the K_{IC} is expected to reduce to a steady value when the density of nodes attains a certain level. Thus, the focal objective is to obtain the optimal number of rings Q acquired for the solutions to converge. The investigation is carried out for 0.4%, 0.5% and 0.6% of crack length a . Up to a total of eight rings are analyzed. The rings are distributed with equal distance between each other in the localized-region.

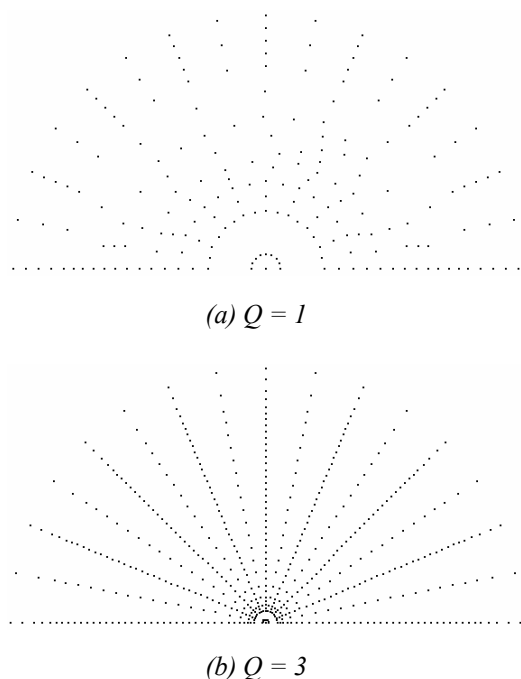


Figure 3. Nodes Distribution for Varying Density Levels

5. Results and Discussions

The numerical results of percentage difference between experimental K_{IC} and numerical K_{CT} for meshing with-localized-region (M-WLR) and without-localized-region (M-WOLR) are tabulated in the Table 1. All degrees of meshing sizes R ranging from 100 to 500microns has shown that M-WLR yielded much lower percentage of variation compared to M-WOLR. The solutions of M-WOLR have a steadier average value of about 31.7%, whereas the solutions of M-WLR are proportionally reduced with diminishing sizes of mesh elements. High degree of mesh refinement at the crack-tip is found to produce the lowest variation up to 5.2%. The best result from the numerical analysis has thus favored the localized-meshing, rather than the large-scale-meshing.

Table 1. Numerical Results for Meshing Techniques

Radius of Elements R (microns)	% Variation	
	M-WLR	M-WOLR
100	5.241	31.757
150	5.281	31.770
200	5.344	31.637
250	5.431	31.804
300	5.422	31.766
350	5.618	31.749
400	7.966	34.165
450	13.361	37.933
500	17.930	41.130

The investigation of effect due to rings distribution within the localized-region allows the illustration of the convergence characteristic of solutions to be modified and studied, contrary to the M-WOLR technique. This characteristic has verified that the variation reached a steady value as expected when a level of nodes density is achieved. In this investigation using three different values of crack length a , the optimal density levels are represented by 5 rings and above. The least density level shows the variation achieved at as low as 4.5%. Numerical data in the tabulated form and graphical representation are shown in the Table 2 and Figure 4, respectively.

Table 2. Numerical Results for Rings Distribution

No. of rings Q	% Variation		
	(i) 0.4% a	(ii) 0.5% a	(iii) 0.6% a
1	10.530	9.544	8.089
2	10.495	9.440	8.031
3	9.820	9.030	7.865
4	5.398	6.392	5.907
5	4.729	5.499	5.054
6	4.692	5.521	4.944
7	4.548	5.472	4.855
8	4.548	5.478	4.855

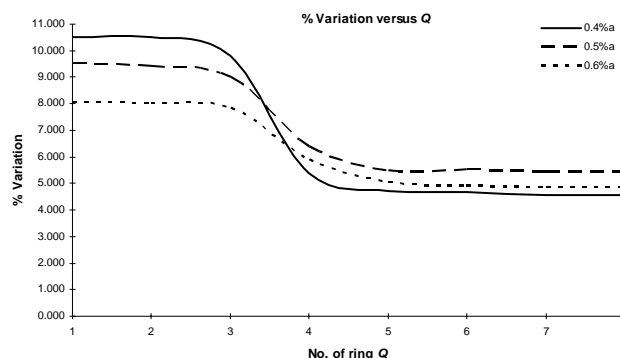


Figure 4. Graphical Representation for Rings Distribution

6. Conclusion

The M-WLR or localized-meshing technique has shown to be a better approach for modeling micro-scale fracture phenomena at the crack-tip region. The investigation has also shown the importance of localized-stress study. The localized parameter of rings distribution within localized-region has been specifically selected to demonstrate the characteristic findings for the purpose of this paper publication. A more complete range of parameters that characterizes the localized crack-tip region can be further studied in the next investigation.

Acknowledgements

The authors would like to acknowledge the Centre of Materials & Minerals of Universiti Malaysia Sabah for the full support of this research project.

References

- [1] Chung DDL. Applied Materials Science: Applications Of Engineering Materials In Structural, Electronics, Thermal, And Other Industries. NY: CRC Press, 2001.
- [2] Sanford RJ. Principles of Fracture Mechanics. NJ: Pearson Education, 2003.
- [3] Collins JA. Failure of Materials In Mechanical Design: Analysis, Prediction, Prevention. 2nd ed. Toronto: John Wiley & Sons, 1993.
- [4] Cotterell, B. 2002. The Past, Present, and Future of Fracture Mechanics. *Engineering Fracture Mechanics*. 69(5): 533-553.
- [5] Courant R. Variational method for the solution of problems of equilibrium and vibrations. Bulletin of the American Mathematical Society. 1943; 49.
- [6] Clough RW. The finite element method in plane stress analysis. Proceedings, American Society of Civil Engineers. Second Conference on Electronic Computation. 1960.
- [7] Araujo TD, Bittencourt TN, Roehl D, Martha LF. Numerical estimation of fracture parameters in elastic and elastic-plastic analysis. European Congress on Computational Methods in Applied Sciences and Engineering. 2000.
- [8] ASTM E399. Standard test method for plane-strain fracture toughness of metallic materials. Annual Book of ASTM Standards. 1990; 3(1): 1–31.
- [9] Fenner RT. Finite element methods for engineers. London: Imperial College Press, 1996.
- [10] Anderson TL. Fracture mechanics: Fundamentals and applications. 2nd ed. Florida: CRC Press Inc, 1995.

Effect of Substrate Bias Voltage on The Surface Characterization of TiN-Coated HSS Using CAPVD Technique

A. Mubarak¹, E. Hamzah^{1,#}, M. R. M. Toff²

¹ Department of Materials Engineering, Faculty of Mechanical Engineering,
Universiti Teknologi Malaysia, 81310 Skudai, Johor, MALAYSIA.
Tel. +60-7-5534653, Fax. +60-7-5576820, e-mails: mubarak60@hotmail.com, esah@fkm.utm.my

² Advanced Materials Research Center (AMREC), SIRIM Berhad,
Tel. +60-4-4017151, Fax. +60-4-4033224, e-mail: mradzit@sirim.my.

ABSTRACT

In the present study, titanium nitride (TiN) coatings on high-speed steel (HSS) were produced using cathodic arc physical vapour deposition technique. We studied and discussed the effect of various substrate biases on the surface properties of TiN-coated HSS. The coating properties investigated includes the surface morphology/topography, surface roughness analysis, and cross-section shape, configuration and fractal dimension analyses. Minimum values for average and root-mean-square roughness were recorded at substrate bias of -150 V, whereas maximum at zero substrate bias. The decrease in surface roughness by increasing substrate bias was mainly associated with the repulsive force, re-sputtering of titanium particles and an increase in substrate temperature. By increasing bias voltage from 0 to -150 V, bearing ratio (%) increased, whereas fractal dimension decreased about 1.24 %. By increasing substrate bias voltage from zero to -150V, about 1.2% decrease in coating thickness was recorded. Macrodroplets were produced during etching stage and protrude through coating. Data recorded via atomic force microscope (AFM) and surface roughness tester show similar trend in their respective data. This scenario promotes an approach toward the optimization of coating properties for specific application.

Keywords: Titanium nitride; Bias voltage; Surface roughness; Morphology; Bearing ratio.

1. Introduction

It is well known that the microstructure of a film, which determines its resultant physical properties, strongly depends on the deposition conditions [1,2]. PVD coatings generally do not have a chemical effect on the substance material composition because of the low temperature process. In addition, it is an environment-friendly process. The bonding in titanium nitride has to be described as a mixture of three types of bonding, metallic, covalent and ionic which simultaneously contribute to the cohesive energy [3].

The inclusion of macroparticles and the formation of pitting defects have been considered detrimental features of cathodic arc evaporated (CAE) coatings since the commercialization of these coatings for cutting-tool and corrosion-protection applications in the early 1980s [4-5]. Over the years various chamber configurations and deposition techniques have been developed to reduce the amount of coating defects [6-8].

The AFM was shown to be a useful tool for studying the surface topography of technologically interesting surfaces. The AFM is a well-suited tool to investigate

* Corresponding Author. E-mail: esah@fkm.utm.my,
Tel: +60-7-5534653, Fax: +60-4-4033224

forces at the nanometer range [9]. Morant *et al.* [10] used AFM to study nanotribological properties of CN/TiCN/TiN/Si multilayer. Glew *et al.* [11] studied the porosity of titanium nitride (TiN) films grown on MgO and silica substrates.

Pre-evaluation of the surface roughness of TiN-coated tool is an important aspect in term of high performance and dry machining. In this study, surface roughness of TiN PVD coated high-speed steel (HSS) were investigated to find out the effect of substrate bias on roughness amplitude parameters, roughness spacing parameters, statistical parameters and bearing ratio analysis. An insight into the surface characterization at micron and nanometric level of TiN PVD coatings will undoubtedly be beneficial to improve the performance of various coated tools.

2. Experimental process

Before charging the samples into vacuum chamber, the base materials were washed in an ultrasonic bath with a mild solvent-free detergent for 30 minutes, rinsed in a stream of cool tap water, dried with blower, and then again blow-dried with high-pressure nitrogen gas to remove any dust and lubricant on the sample's surface.

TiN single layer coatings have been deposited in a commercially available Hauzer Techno Coating (HTC) 625/2 ARC coating system, a detailed description of which is given elsewhere [12]. It has been demonstrated that the metal ion bombardment favors local epitaxial film growth, which enhances adhesion between coating & substrate [13]. During deposition, the substrates were biased with a dc power source to induce proper ion bombardment on the growing surface to assist the formation of desirable structure, grain size and film density. It has been shown that the direct current implantation is more effective than the pulsed implantation in the surface corrosion resistance [14].

For the deposition of the TiN coatings, a constant amount of nitrogen gas was introduced into the coating chamber at arcs current of 100 A during each experiment. The objective of this present work is to see, how substrate bias voltage behaves on the surface roughness parameters. This is demonstrated using single-phase TiN coatings deposited on HSS by using CAPVD technique.

3. Results and Discussion

3.1. Atomic force microscopy (AFM) studies

The AFM operates by positioning a very small lever (cantilever) close to the sample surface and detects the force (atomic force) between the cantilever and sample surface to allow the features of the sample surface to be observed. Generally, the Van der Waals force and Coulomb force act on the surfaces of materials. All the scans were undertaken in the atmosphere without pretreatment such as coating i.e. in case for SEM observations. The cantilever end can be moved within an area of 4.0×4.0 mm with respect to a sample by adjusting the observation point alignment stage of the AFM head unit. Prior to the investigations of TiN coatings, the excellent performance of the AFM was confirmed by imaging a standard sample for calibration provided by the manufacturer.

3.1.1. Surface morphology

The topography of the surface samples was observed using the scanning probe microscope (model SPM-9500J2, Shimadzu Corporation) with an AFM that was controlled via the Microsoft Window 95 OS. Figure 1 shows the three-dimensional (3D) images of the TiN coating deposited at zero and -150 V substrate biases. It has been reported that the difference between the structures is attributed to differences in the preferential growth direction and anisotropy [15]. There is a strong dependence of surface roughness upon surface topography [16]. The AFM images in Figures 1 (a) and 1 (b) reveal the 3D surface topographic feature of TiN-coated HSS. At zero bias the grains were formed in peaks and valleys-island shaped particles and can be identified. The peaks represent the macrodroplets, resulted from the melted titanium cathode/target. As the substrate bias increased to -150 V, reduction in macrodroplets occurred because of the repulsive force generated between high negative substrate bias and charged particles just after arrival on substrate surface. Most of the bigger size particles treated neutral before the arrival on substrate surface and after the ejection from cathode. It

means, as the titanium particles from the sheath and plasma stream occupy the substrate surface, they were not fully charged. Once these particles occupied the substrate surface, they become charged particles because of the penetration and bombardment of charged particles from the plasma stream. Substrate temperature was raised because of increased substrate bias voltage. In these circumstances, these particles (macrodroplets) re-sputtered and the result was dark and hollow spot on the end coated surface and can be seen in Figure 1 (b). These macrodroplets later on evacuated from the sheath of vacuum chamber.

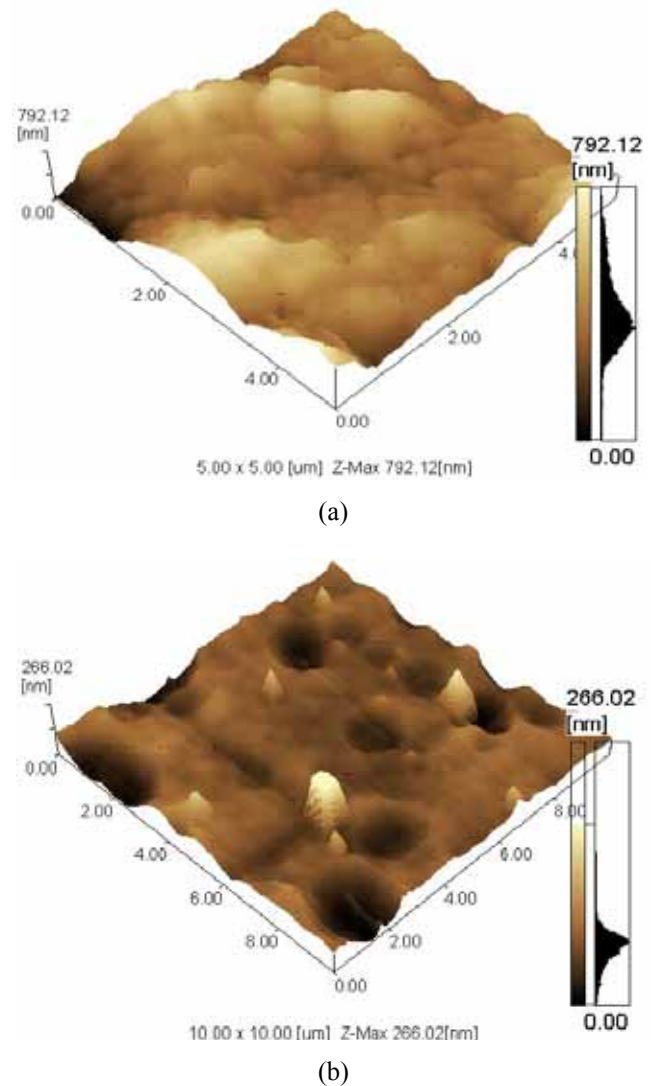
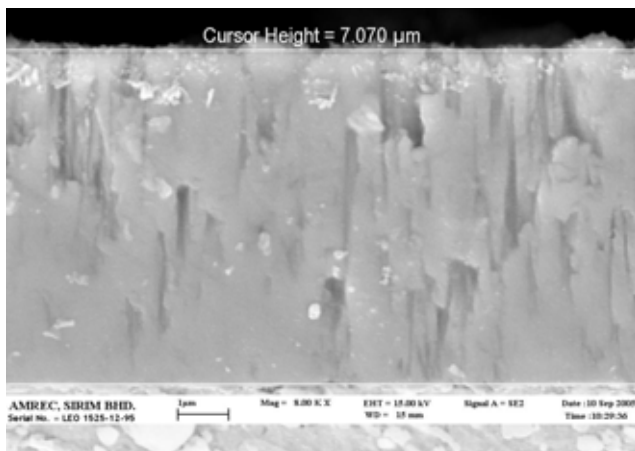


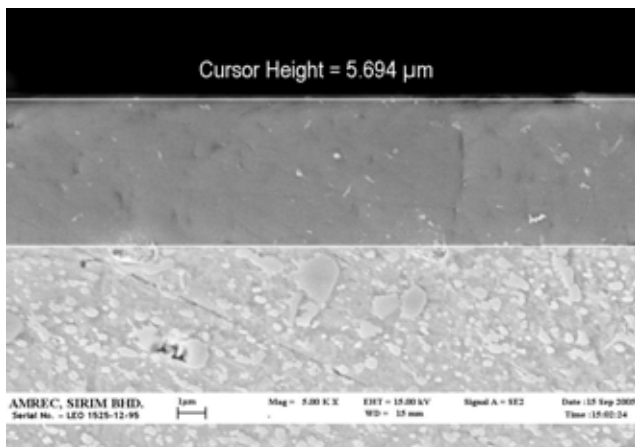
Figure 1. AFM three-dimensional (3D) images at bias voltage (a) 0 and (b) -150 V.

As it stands, that under high negative substrate bias, the temperature of substrate increased. Therefore re-sputtering occurred because of negatively charged substrate with high temperature and charged particles. This mechanism was observed during the coating stage. During the etching/pretreatment stage, an interlayer of titanium under pre-bias of -1000 V for 5 minutes in the presence of Argon gas was grown. Etching conditions were the same in all coatings and the intention was basically to improve the adhesion strength between the substrate and TiN coating.

Our previous work shows that most of the macrodroplets were produced during the etching stage instead of the coating stage [17]. Figure 1 (b) also indicates, some of the nail head like particles. These nail-head like particles were produced during etching stage in the form of macrodroplets. The nail-head form of these droplets resulted because of the severe attack of high energy fast moving electrons coming from the plasma stream, exactly perpendicular to them. These macrodroplets produced during the etching stage have their stems into the coating and later on protrude through the coating. The existence of these macrodroplets in the coating create serious problem during high-speed machining, dry machining and high performance cutting and ultimately affect the performance of coated tools e.g. during drilling, milling, turning and etc.



(a)



(b)

Figure 2. SEM micrographs showing thickness of deposited coatings at (a) zero bias and (b) -150 V.

It is interesting to note here, that the scanning electron microscope is not an appropriate technique to study the nucleation and growth mechanisms and therefore transmission electron microscopy normally operates to tackle this matter. AFM used in the present study not only provides us the information about surface morphology

(without any surface pretreatment) and surface roughness parameters but also provide information about nucleation and growth mechanisms. From the above discussion, it can be said that AFM is a powerful technique for surface characterization, whose importance relies on its resolution power, allowing the acquisition of topographic images of the surface with nanometric resolution.

By increasing the substrate bias voltage also decreased the thickness of deposited coatings which can be seen in Figure 2. The deposited coatings de-attach at higher substrate biases because of increase in substrate temperature and re-sputtering of macroparticles.

3.1.2. Surface roughness analysis

The AFM analysis provided us the surface roughness as a function of the lateral length scale and an estimation of the columnar grain size. The AFM is also suitable for quantitative study of surface roughness and scaling parameters [18]. The following surface roughness parameters were studied during AFM studies:

- **R_a**, average roughness, also known as arithmetic average/area average (AA), centre line average (CLA), arithmetical mean deviation of the profile and calculated as the area between the roughness profile and its mean line and is given by equation (1)

$$R_a = \frac{1}{N} \sum_{i=1}^N |Y_i| \quad (1)$$

R_a is the arithmetic mean of the absolute values of the profile deviations (Y_i) from the mean line. The main limitation of R_a is that it gives no indication of the surface texture, namely macrodroplets and pitting defects that could have a significant influence on the performance of the coating, particularly for wear resistance applications. The average roughness is by far the most commonly used parameter in surface finish measurement.

- **R_y**, is the maximum height of the profile or total roughness (R_t). R_y is the sum of height Y_p of the highest peak from the mean line and depth Y_v of the deepest valley from the mean line and is given by equation (2)

$$R_y = Y_p + Y_v \quad (2)$$

- **R_z**, indicates the average maximum height of the roughness profile. $R_z = R_p + R_v$ and is the maximum peak to valley height of the profile in a single sample length. R_z (JIS) -also known as the ISO 10-point height parameter, and is the sum of the mean height of the five highest profile peaks and the mean depth of five deepest profile valleys measured from a line parallel to the mean line and is given by equation (3)

$$R_z \text{ (JIS)} = \frac{1}{5} \sum_{i=1}^5 Y_{pi} + \frac{1}{5} \sum_{i=1}^5 Y_{vi} \quad (3)$$

- **R_{ms}**, Root-mean-square deviation of the profile, also denoted by R_q . R_q is the square root of the arithmetic mean of the square of profile deviations (Y_i) from the mean line and is given by equation (4)

$$Rms = \left(\frac{1}{N} \sum_{i=1}^N Y_i^2 \right)^{1/2} \quad (4)$$

- **Rp**, is the maximum profile height; first obtain the profile peak height *Rpi* for each sampling length of the assessed profile. The mean of the *Rpi*'s obtained over the evaluation length is the *Rp* and given by equation (5)

$$Rp = \frac{Rp1 + Rp2 + Rp3 + Rp4 + Rp5}{5} \quad (5)$$

- **Rv**, is the maximum profile valley depth; first obtain the profile valley depth *Rvi* for each sampling length of the assessed profile. The mean of the *Rvi*'s obtained over the evaluation length is the *Rv* and is given by equation (6)

$$Rv = \frac{Rv1 + Rv2 + Rv3 + Rv4 + Rv5}{5} \quad (6)$$

Ra and *Rms* decreased by increasing substrate bias voltage and can be seen in Table 1. A large difference in the surface roughness was recorded for the coating deposited at substrate bias of zero and -150 V.

3.1.3. Configuration analysis

Configuration analysis at substrate bias voltage of -150 V can be seen in Figure 3. The data obtained from configuration analysis at various substrate biases can be seen in Table 1. From Table 1, it was noted that the bearing ratio (%) increased with an increase in substrate bias

voltage, whereas, the maximum value for histogram (%) was recorded at zero substrate bias. Surface area and volume of the selected region decreased with an increase in substrate bias voltage. Whereas, the maximum value for height was noted at zero bias and minimum at -150 V.

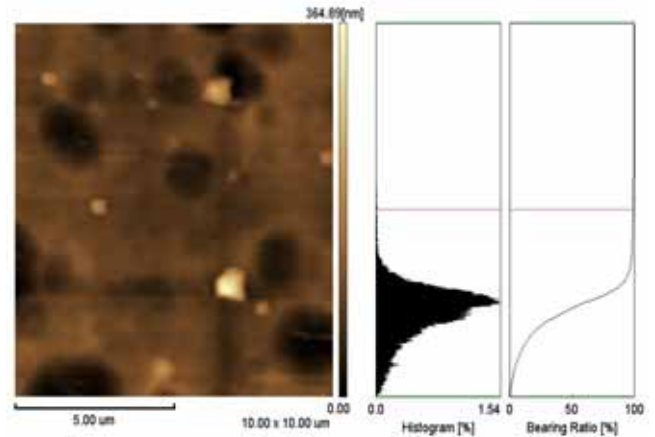


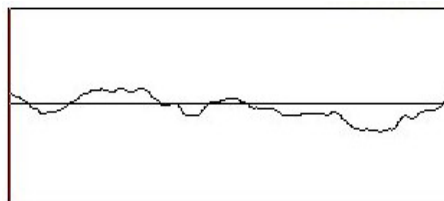
Figure 3. Configuration analysis of TiN coating at bias voltages (a) 0 and (b) -150 V.

The bearing ratio curve mathematically is the integral of the amplitude distribution function. It is a cumulative probability distribution. Ordinary, the integral is performed from the highest peak downward, so each point on the bearing ratio curve has the physical significance of showing what linear fraction of a profile lies above a certain height.

Table 1. AFM data of TiN-coated HSS at various substrate biases.

Bias voltage (V)		0	-25	-50	-100	-150
Surface roughness analysis	Length-X (µm)	10.000	10.000	10.000	10.000	10.000
	Length-Y (µm)	10.000	10.000	10.000	10.000	10.000
	Area (µm) ²	100.00	100.00	100.00	100.00	100.00
	Ra (nm)	217.874	138.954	80.216	40.541	20.637
	Ry (nm)	2258.0	950.605	964.016	561.070	365.116
	Rz (nm)	1291.0	651.918	401.954	357.454	319.349
	Rms (nm)	286.841	174.557	118.934	55.268	29.943
	Rp (nm)	1096.0	603.407	638.560	341.603	279.704
Configuration information	Rv (nm)	1162.0	347.198	325.455	219.467	85.413
	Range Cursor (nm)	2260.0-0	949.84-0	963.96-0	561.58-0	364.89-0
	Surface area (µm) ²	116.527	110.788	106.697	102.312	100.811
	Volume (µm) ³	116.514	34.9434	32.7364	22.0978	8.53575
	Height (nm)	1130	475.85	482.92	281.34	182.80
	Histogram (%)	0.66	0.24	0.11	0.34	0.00
Line profile analysis	Bearing ratio (%)	48.66	81.24	91.25	91.58	93.13
	Width (µm)	≈14.0	≈14.0	≈14.0	≈14.0	≈14.0
	Ra (nm)	260.12	120.06	*	33.84	18.56
	Ry (nm)	986.43	799.86	*	181.49	98.37
	Rz (nm)	497.94(8)	368.42(10)	*	123.88(8)	46.3(10)
	Sm (µm)	1.20	2.61	*	3.29	2.63
	S (µm)	0.66	0.41	*	0.51	0.54
Fractal dimension	Rms (nm)	299.60	164.92	*	40.62	22.39
		2.03733	2.03863	*	2.00130	2.00038

*Data not available



exponent of the number of self-similar pieces with magnification factor N into which the figure may be broken. For the square, we have $N = 2$ self-similar pieces, each with magnification factor N . So we can write

$$\text{Fractal dimension} = \frac{\log(\text{number of self-similar pieces})}{\log(\text{magnification factor})}$$

Figure 4. Line profile analysis of 2D image via AFM.

3.1.4. Cross-section shape analysis (Line profile)

A line profile of two-dimensional (2D) image of can be seen in Figure 4. A quantitative data obtained during this line profile analysis via AFM can be seen in Table 1. From Table 1, it was noted that the Ra and Rms show similar trend, which was noted during surface roughness analysis. The following line profile parameters were studied during AFM studies:

- S_m , is the mean width/spacing of the profile elements (irregularities). A portion projecting upward over the given upper count level is called a profile peak, and a portion projecting downward below the given lower count level is called a profile valley. The width of the profile element is the length of the x-segment intersecting with profile peak and the adjacent profile valley. S_m is the mean of the profile element widths within a sampling length and is given by equation (7)

$$S_m = \frac{l}{n} \sum_{i=1}^n |S_{mi}| \tag{7}$$

- S , is the mean width/spacing of local peaks of the profile. S is the arithmetic mean of peak-to-peak distances of the local peaks and is given by equation (8)

$$S = \frac{l}{n} \sum_{i=1}^n |S_i| \tag{8}$$

3.1.5. Fractal dimension analysis

Fractal dimension is a measure of how "complicated" a self-similar figure is. In a rough sense, it measures "how many points" lie in a given set. The dimension is simply the

$$= \frac{\log N^2}{\log N}, = \frac{2\log N}{\log N} = 2$$

Similarly, the dimension of a cube is 3.

The maximum value of fractal dimension was recorded at -25 V, whereas the minimum is at -150 V. This means that the size of number of self-similar pieces in TiN coating decreased by increasing bias voltage and vice versa.

3.2. Surface roughness tester

The results of quantitative surface roughness tests on uncoated and coated HSS coupons are reported in Table 2. Surface profiles for the coatings deposited at zero and -100 V bias voltages can be seen in Figure 5. Harris *et al.* [19] calculated the roughness of the TiN coatings on HSS twist drill, and found that it increased from 0.104 to 0.116 μm as the chamber pressure decreased from 1.2 to 0.1 Pa.

Table 2. Surface roughness at various substrate biases.

Substrate bias voltage	Surface roughness (ISO 97 Gauss),		
	Ra (μm)	Rz (μm)	Rq (μm)
Uncoated	0.02	0.23	0.03
0	0.43	4.04	0.61
-25 V	0.35	4.40	0.54
-50 V	0.27	3.26	0.44
-100 V	0.21	3.71	0.41
-150 V	0.18	2.55	0.32

The surface roughness (Ra) of uncoated HSS was about 0.02 μm . The maximum Ra value was noted for coating deposited at zero bias voltage, whereas, the minimum value for the coating deposited at -150 V.



Figure 5. Surface profiles of CAE PVD coatings at -100 V ($EVA-L = 4.0 \text{ mm}$, $\lambda c = 0.8 \text{ mm} \times 5$).

The increase or decrease in surface roughness of TiN coating depends on the PVD process character-cathodic arc evaporation and occurrence of the characteristic macro droplets, due to deposition of the pure titanium droplets, coming from the sputtered target and therefore, pits developing due to the titanium micro-particles dropping

out, immediately after finishing the coating deposition process. Therefore, the surface of the coating demonstrate inhomogeneities connected with shaped occurrences of the droplet-shaped, elongated micro-particles and growth defects originating probably from their sputtering when

they hit the substrate surface during the coating deposition process.

4. Conclusions

All coatings were deposited under the same deposition parameters except substrate bias voltage. The decrease in fractal dimension with an increase in bias voltage, basically indicate reduction in pinholes and hence decrease in intrinsic stress of films. It was noted that R_a and R_{ms} of TiN-coated HSS decrease by increasing substrate bias. The measured parameters as R_a , R_q and R_z from both AFM and surface roughness tester show similar trend in their respective data. A similar trend is noted in R_a and R_{ms} during surface roughness analysis and cross-section shape analysis. The decrease in surface roughness by increasing bias voltage is mainly associated with increase in substrate temperature and repulsive force, and hence reduction in both the size and amount of macrodroplets. Reduction in macrodroplets mainly contributes to a decrease in the surface roughness. By increasing substrate bias decreases the coating thickness and hence compressive stress. This decrease in compressive stress also helps in decreasing the surface roughness of TiN coatings. The macrodroplets have there stems in the base material and lay through the TiN film.

Acknowledgments

Thanks are due to the Government of Malaysia for the scholarship for Mubarak Ali and the Malaysian Ministry of Science, Technology and Innovation from the IRPA funding Project No: 03-02-06-0067 PR0006/06-02 and to AMREC SIRIM Berhad for the facilities and technical support.

References

- [1] Greene, J. E., Barnett, S. A., Sundgren, J. - E. and Rockett, A. eds. 1989. *Ion Beam Assisted Film Growth (Chapter 5)*. Amsterdam: Elsevier, p. 101.
- [2] Martin, P. J., Netterfield, R. P. and Sainy, W. G. 1984. Modification of the optical and structural properties of dielectric ZrO_2 films by ion-assisted deposition. *Journal of Applied Physics* 55: 235-241.
- [3] Blaha, P. and Schwarz, K. 1983. Electron densities and chemical bonding in TiC, TiN, and TiO derived from energy band calculations. *Int J Quantum Chem* 23: 1535-1552.
- [4] Matthews, A. and Lefkow, A. R. 1985. Problems in the physical vapor deposition of titanium nitride. *Thin Solid Films* 126: 283-291.
- [5] Disatnic, G., Boxman, R. L. and Goldsmith, S. 1987. Characteristics of Macroparticle Emission from a high-Current-Density Multi-Cathode Spot Vacuum Arc. *IEEE Trans. Plasma Sci.* PS-15: 520-523.
- [6] Coll, B. F. and Sanders, D. M. 1996. Design of vacuum arc-based sources. *Surf. Coat. Technol.* 81: 42-52.
- [7] Sanders, D. M. and Anders, A. 2000. Review of cathodic arc deposition technology at the start of the new millennium. *Surf Coat Technol.* 133-134: 78-90.
- [8] Martin, P. J. and Bendavid, A. 2001. Review of the filtered vacuum arc process and materials deposition. *Thin Solid Films* 394: 1-15.
- [9] Burnham, N. A. and Colton, R. J. 1989. Measuring the nanomechanical properties and surface forces of materials using an atomic force microscope. *J Vac Sci Technol A* 7: 2906-2913.
- [10] Morant, C., Garcia-Manyes, S., Sanz, F., Sanz, J. M. and Elizalde, E. 2005. Nanotribological properties of CN/TiCN/TiN/Si multilayer as determined by AFM. *Nanotechnology* 16: S211-S217.
- [11] Glew, M. R. L., Vollmer, A., Schroeder, S. L. M. and Barber, Z. H. 2002. The characterization of TiN thin films using optical reflectivity measurements. *J Phys D: Appl. Phys.* 35: 2643-2647.
- [12] Mubarak, A., Hamzah, E., Toff, M. R. M. and Hashim, A. H. 2005. The effect of Nitrogen gas flow rate on the properties of TiN-coated HSS using Cathodic Arc Evaporation PVD Technique. *Surface Review and Letters* 12: 631-643.
- [13] Petrov, I., Losbichler, P., Bergstrom, D., Greene, J. E., Münz, W. D., Hurkmans T. and Trinh, T. 1997. Ion-assisted growth of $Ti_{1-x}Al_xN/Ti_{1-y}Nb_yN$ multilayers by combined cathodic-arc/magnetron-sputter deposition. *Thin Solid Films* 302: 179-192.
- [14] Zhang Gu-Ling, Wang Jiu-Li, Liu Yuan-Fu, Liu Chi-Zi and Yang Si-Ze. 2004. Properties of TiN coating on 45#steel for inner surface modification by grid-enhanced plasma source ion implantation method. *Chinese Physics* 13: 1309-1314.
- [15] Westra, K. L. and Thomson, D. J. 1995. The microstructure of thin films observed using atomic force microscopy. *Thin Solid Films* 257: 15-21.
- [16] Martan, J., Przybylski, G., Tabaka, R. And Kowalski, Z. W. 2005. Fractal analysis of roughness profile induced by ion bombardment of metal surface. *Vacuum* 78: 217-221.
- [17] Mubarak, A., Hamzah, E. and Toff, M. R. M. The effect of metal ion etching on the properties of TiN-coated high-speed steel (HSS) using cathodic arc evaporation physical vapor deposition (PVD) technique, *Brazilian Journal of Physics* (under review process).
- [18] Krim, J. and Palasantzas, G. 1995. Experimental observations of self-affine scaling and kinetic roughening at sub-micron length scales. *Int J Mod Phys B* 9: 599-632.
- [19] Harris, S. G., Doyle, E. D., Wong, Y. C., Munroe, P. R., Cairney, J. M. and Long, J. M. 2004. Reducing the macroparticle content of cathodic arc evaporated TiN coatings. *Surf. Coat. Technol.* 183: 283-294.

Strain-Induced recrystallization kinetics of Nb-high strength low alloy steel during hot rolling

E.S.Siradj

Department Metallurgy and Materials
Engineering Faculty University of Indonesia.
Kampus UI Depok 16424, Tel(021)7863510 Fax: 021 7872350
Email: siradj@metal.ui.ac.id

Abstract

During thermo-mechanical treatment process of Nb- High Strength Low Alloy Steel, the grain size of final microstructure ferrite after transforming from austenite grains has influenced on mechanical properties. It means that the smallest austenite grains, better mechanical properties will be achieved due to finest-formed ferrite grains. During hot deformation, temperature and strain have a significant influenced on the size of recrystallized austenite grains. Using HSLA steel containing 0.031% Nb and hot rolling method, the relationship between the size of prior and recrystallized austenite grain, temperature and strain were studied. The established empirical model was evaluated and validated for predicting on this steel. The results show that the empirical model should be modified by considering the strengthening of this steel.

INTRODUCTION

The evaluation of strain induced recrystallisation of High Strength Low Alloy steel containing 0,03%Nb is part of progress report which has been done by post graduate research group Thermomechanical Control Process in Metallurgy and Matarials Department Engineering Faculty University of Indonesia.

The development of microalloyed steels, including their alloy design, processing and applications, covers the last four decades[2]. During this period, microalloyed high-strength, low alloy (HSLA) steels became a indispensable class of structural steels.

The role of high strength low alloyed steel HSLA will become important in the future. M.Korchynsky [2] in his paper a new role for microlloyed steel adding economic value, said that despite these improvements, the total consumption of microalloyed steel is currently estimated to be only 10 to 15% of the world's steel production (i.e. 80 to 120 million tons per year). As a result, there is plenty of room for growth. A major jump in the usage of microalloyed steels should have strong economic benefits for both steel producers and steel users.

Extensive research work has been carried out on the microstructural changes produced by hot working condition at high temperature. Kwon [3] has showed that microstructural evaluation of C-Mn steel and High strength low alloyed steel in term of mathematic empirical model has been carried out by several researcher. The model described the solubility temperature of Nb(CN) precipitate, prior and recrystallised austenite grain growth, fraction recrystallised of austenite grain, strain induced of precipitation kinetics, transformation of austenite to ferrite phase and mechanical properties. Most of the established model is derived from plane compressive test spacemen.

The model for static recrystallisation kinetics in high strength low alloy steel before the start strain induced precipitation has been proposed by Sellar[4] based on the modified avrami equation can be written as,

$$X = 1 - \exp[-0,693 (t_x/t_{0,5})^n] \quad (1)$$

Where,

X is farction recrystallised
 t_x is time for X fraction recrystallised
n is constant with the value of 2[4] and 1,3[5]
 $t_{0,5}$ is time for 50% recrystallised.

Time for 50% recrystallisation, $t_{0,5}$, has been modeled [4] as

$$t_{0,5} = A \varepsilon^p d_0^q Z^r \exp(Q_{rsx}/RT) \quad (2)$$

Here,

d_0 is initial grain size (μm)
Z is Zener-Hollomon parameter
($Z = \varepsilon \exp(Q_{rex}/RT)$)
 Q_{rex} is apparent activation energy for recrystallisation (J/mol)
 Q_{def} is apparent activation energy for deformation (J/mol)

A,p,q and r are material dependent constant.

It is known that microalloying element such as Nb, remarkably influence microstructure of steels through precipitation of their carbonitrides even though their addition is usually limited to 0,01 to 0,1 wt%. Dutta and Sellars [1] have proposed a model for 50% recrystallisation of austenite grains when all Nb is in solution,

$$t_{0,5} = [-5.25 + 550(\text{Nb})] \times 10^{1.8} \varepsilon^{-4.0+7.7(\text{Nb})} \text{do}^2 \exp(330.000/\text{RT}). \quad (3)$$

While Medina[6] suggested that time for 50% recrystallisation for Nb-steel containing 0,042%Nb, 0,11%C, as given by,

$$t_{0,5} = 3,68 \times 10^{1.2} \varepsilon^{1.9} \varepsilon^{-0.44} \text{do} \exp(330.000/\text{RT}). \quad (4)$$

In order to evaluate the recrystallised grain size, a empirical relations based on observations in single deformation studies of static recrystallisation has proposed by Sellar[4]. The model can be rewritten as,

$$d_{\text{rex}} = A \text{do}^{0.67} \varepsilon^{-0.67} \quad (5)$$

Where
do is initial austenite grain size
 ε is applied strain
A is constant

Cuddy [7] has found that $A = 0,75$, while Valdes and Sellars [5] obtained $A = 0,9$ and Liu [8] established $A = 1,1$.

Following complete static recrystallisation, the equiaxed austenite microstructure can coarsen by grain growth. Beck et al[9], firstly showed that experimentally determined value of grain sizes during normal grain growth under isothermal annealing condition fitted a power relation as

$$d^n - \text{do}^n = CT \quad (6)$$

Where d is the final grain diameter, do in the initial grain diameter, t is the annealing time, n and C are constants which depend on alloy composition and annealing temperature. Later, Sellar and Whitemen [10] suggested that the grain growth model, under isothermal condition can be described as,

$$d^n - \text{do}^n = [A \cdot \exp(-Q_{\text{gg}}/\text{RT})] \cdot t \quad (7)$$

where n and A are constants which depend on material composition and processing condition, Q_{gg} is the activation energy for grain growth. For the most metals and alloy the value of n varies typically in the range from 10 to 4 [11] due to the drag force exerted by impurity elements in solid solution. The value of n for steel containing micro alloying such Nb is obtained of 4,5 [1,12]. The activation energy for grain growth, Q_{gg} and the constant A are 435 kJ/mol and $4,1 \times 10^{12}$.

Even though the model of austenite grain growth is formulated for prediction of austenite grain growth after deformation (recrystallisation), several researcher, Manohar et al [13] has attempted to evaluate the grain growth model for austenite grains during reheating. The results show that the empirical models for grain growth are

inadequate for quantitative description of austenite grain growth during reheating for microalloyed steels. However, several models has been established such as Komatsubare et al [14] based on the normal growth rate proposed by Hillert[15].

In the present paper an attempt is made to evaluate the effect of applied strain on the austenite grain growth through hot rolling rather than using hot plane compressive strain. High strength low alloy steel containing 0,039 wt % received as structural steel is used and the the recrystallised grain size model is evaluated.

MATERIAL AND EXPERIMENTAL PROCEDURE.

A high strength low alloy steel containing single micro alloying of 0,038% wt % Nb as received as structural steel is considerably used in this experiment. The chemical composition of those steel is listed in table 1 below,

Table 1. Chemical composition of steel investigated, wt %

C	Si	Mn	P	S	Al	N	Nb
0.1	0.37	1.35	0.02	0.019	0.041	0.0042	0.031

Specimens were subsequently machined from structural hot rolled plate, as illustrated in Figure 1. The surface of specimen were chromium plated to 15 μm thickness to prevent oxidation and scaling during reheating and rolling.

The specimens were firstly reheated by using carbolite muffle furnace at a temperature of 100° C above the solubility temperature. The solubity temperature of Nb(CN) was calculated by using empirical model proposed by Irvine et al [16] as

$$\text{Log}[\text{Nb}][\text{C}+12/14\text{N}] = 2.26 - 6770/\text{T}$$

Here, [Nb] and [C+12/14N] are describing wt % of Nb, C and N, while T is dissolution temperature of Nb(CN) in °K. The result of calculation solution temperature leading to a reheat temperature of 1200°C.

After reheating at a temperature of 1200 °C for 30 minutes, the specimens is air-cooled to rolling temperature of 1150 C. The heated specimen is then rolled under strain deformation of 0,3, 0,35, 0,42 and 0,5 by using UNO rolling machine capacity of 20 ton. After rolling, deformed specimen is then air-cooled at holding time of 20 second before quenching. In order to evaluate the effect of roughing strain on austenite grains growth after finishing temperature, deformation strain of 0,35 is carried out at temperature of 1150 °C and then finishing strain of 0,5 at temperature of 1020 °C.

During experiment, the historical temperatures of specimen is recorded through thermocouple wire by using data acquisition and control software.

The effects of temperature and reduction on the recrystallised austenite grain size were studied after the specimens were quenching from finishing temperature. The sample close to the tip of thermocouple wire was cut for metallography examination. To reveal the austenite grain boundaries on the surface of specimen, etching was carried out by using aqueous picric acid containing a small amount of HCL and wetting agent teepol according to procedure of Baldinger et al [17].

The grain size measurements were carried out using the mean linear intercept method. An optical microscope with a micrometer stage was used to measure at least 200 grains in each of the longitudinal and transversal directions.

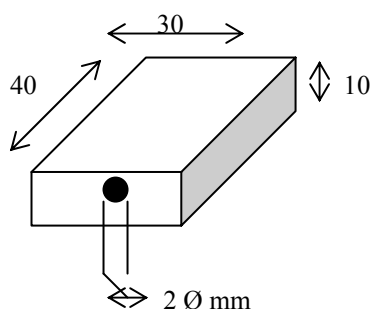


Figure 1. Illustration of test specimen

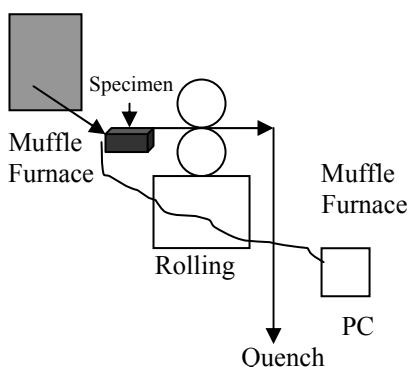


Figure 2. Illustration reheating, rolling and quenching test specimen.

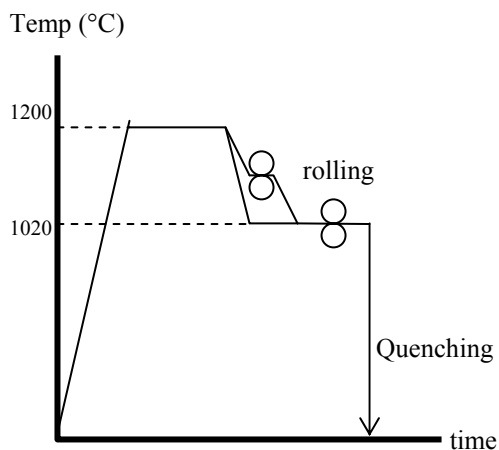


Figure 3. Thermal cycle of tested steel

EXPERIMENTAL RESULTS AND DISCUSSION.

From metallography observation, the prior austenite grain size after reheating at temperature of 1200 °C for 30 minutes is observed around 150 μm. The austenite grain growth behavior of these steel compared with the results obtained by other investigaor can be seen in Figure 4.

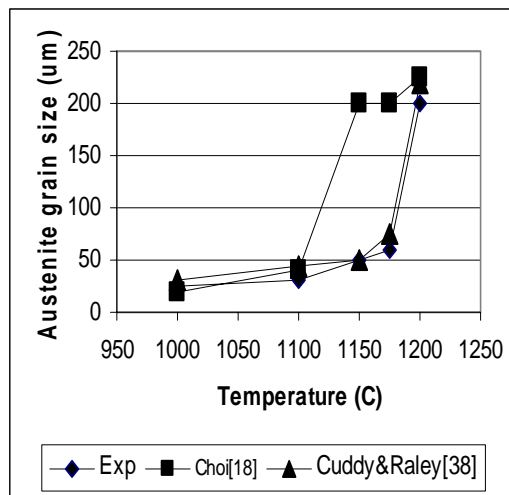


Figure 4. The austenite grain growth behaviour of test steel.

The austenite grain growth of test steel seems to be close agreement with Cuddy and Raley [19] steel. The chemical composition of these steel such as Carbon is quite similar, however for Nb 50% higher then test steel. The prior austenite grain boundaries of test steel after reheating at temperature of 1200 °C for 30 minutes can be seen in Figure 5.

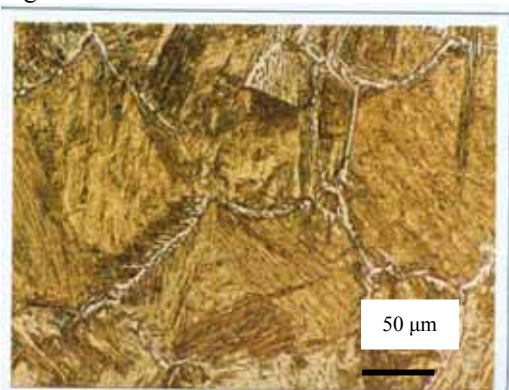


Figure 5 Showing the prior austnite grain boundaries of test steel after reheating at temperature of 1200 C.

The figure 4 shows that the test steel containing 0.031% Nb at temperature up to 1150 °C, the austenite grain size is grow slowly due to pinning effect of the particle Nb(CN). The abnormal austenite grain growth is obtained at temperature around of 1175 °C and above this temperature the austenite grain is growing normally due to Nb(CN) particle has soluble into matrix.

The recrystallised austenite grain is observed after deformation and quenching. Table 2 shows the test condition and austenite grains size.

Table 2. Test condition and austenite grain size results.

Roughing Temp °C	Strain	Finishing Temp °C	Strain	Grain size μm
1150	0.35	1020	0.30	80±8
			0.35	72±8
			0.42	65±7
			0.51	50±5
	0.35	1020	0.30	35±4
			0.35	25±3
			0.42	21±2
			0.51	15±2

The austenite grain size displayed in Table 2 is the grain size after recrystallisation and growth. To evaluate the actual recrystallised grain or 95% fraction of recrystallisation, time for 95% fraction recrystallised should be calculated by using empirical model proposed by Dutta and Sellar in equation (3), (1) and (7). The result can be seen in Table 3.

Table 3 Calculation result of time for 95% recrystallisation, using equation (3),(1) and(7)

Condition	Austenite grain size after quench (μm)	Time for 95% (sec)	Austenite grain size for 95% (μm)
No Roughing	80±8	7.98	79
Roughing	72±8	4.30	71
	65±7	2.07	64
	50±5	1.03	44
Roughing	45±4	0.84	44
	40±3	0.45	39
	39±2	0.22	38
	35±2	0.13	34

To evaluate austenite grain size when roughing deformation is performed, initial austenite grain size before finishing deformation should be calculated using similar equation plus equation (5). The calculation result gives the initial austenite grain size around 65 μm.

The relationship between recrystallised austenite grain size and deformation strain at temperature of 1020°C as displayed in Table 3 can be seen in Figure 6.

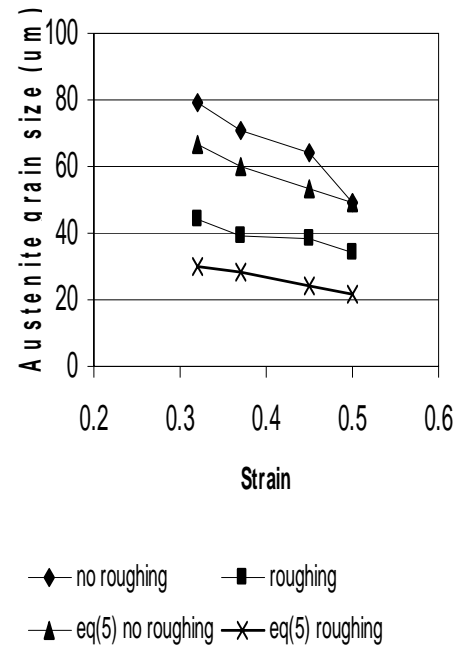


Figure 6. The relationship between strain deformation and austenite grain size compared with established model.

Figure 6, shows that empirical model proposed by Sellar [1] in equation (5) seems not consistence with the experimental result. The model gives the value of recrystallised austenite grains size lower then the actual result in both no roughing and roughing deformation is given. It could be explain that the established model (5) is generated from plane strain compression test result rather than plane strain in hot rolling deformation. The recrystallised austenite grain model (5) could be applied when austenite grain growth in condition of isothermal, whereas in this experiment after rolling at temperature 1020 °C before quenching the temperature of deformed specimen drops about 50 °C. Recrystallisation kinetics of austenite grain in the test steel at low temperature is lower than at a high temperature.

The other explanation could be arising from the distribution of strain. The strain distribution cross section of steel which is deformed by rolling regime is more complex compared with plane compressive test [20]

Looking at the Figure 6, the slope of the established model by Dutta and Sellar seems to be consistence with the experimental results. The model could be used to predict the static recrystallisation kinetics of high strength low alloy steel containing single alloying Niobium (Nb) using hot rolling deformation, when the constant A in the model should be modified from 0.9 to 1.06 for without roughing and 1.3 for roughing. The modified model can be rewritten as,

$$d_{rex} = 1.06 d_0^{0.67} \epsilon^{-0.67} \text{ single pass}$$

$$d_{rex} = 1.3 d_0^{0.67} \epsilon^{-0.67} \text{ double pass}$$

CONCLUSION.

1. Empirical model to predict static recrystallisation in high strength low alloy steel has been proposed by several investigator including model for austenite grain growth behavior during reheating, dissolution of Nb(CN) and recrystallisation kinetic.

2. Most of the model are derived from plane compression and torsion test data, where the strain distribution is less complex than the strain distribution in test steel using hot rolling

3. Established model to predict recrystallised austenite grain function of prior austenite grain and strain deformation seems not close agreement with the experimental data. The model should be modified to predict static recrystallisation kinetics in hot rolling.

REFERENCES

1. B.Dutta and C.M.Sellars;(1987), Mat Sci and Technol, Vol 3.
2. M.Korchynsky (2001), Stratcor, Presented at infacon 9 Quebec City, Canada, June 6.
3. O.Kwon , (1992), ISIJ Int Vo; 32 No 3.
4. C.M.Sellar, (1980), Hot working and forming processes, ed C.M.Sellar and G.J.Davies, London, The Metal Society.
5. E.Valdes and C.M.Sellars (1991), Met Sci and Technol, Vol 7
6. S.F. Madina;(1995),Scripta Met, Vol 32, no 1.
7. L.J. Cuddy, (1982), Thermomechanical Processing of Microalloyed austenite; ed P.J.Wray. H.J.DeArdo, G.A.Rats, Warradale,Pa, The Metallurgical Soc of AIME.
8. W.J. Liu (1995), Metall Trans A, Vol 26A.
9. P.A.Beck and P.R.Sperry (1949), Trans, AIME, Vol 185
10. C.M. Sellars and J.A Whiteman, (1979), Met Sci, vol 13.
11. T.Andesen and O.Grong, (1995), Acta Met, Vol 43 no 7.
12. J.H.baynon and C,M,Sellars, (1992), ISIJ Int, vol 32.
13. P.A.Manohar, D.P.Donne, T.Chandra and C.R.Killmore, (1996), Trans, AIME.
14. N.Komatsubare, K.Kunishige, S.Okegushi, T.Hashimoto, K.Okshime and T.Tamura, (1990), The Sumitomo Search, no 44.
15. M.M.Hillert (1965), Acta Met, Vol 13.
16. K.J.Irvin, F.B.Pickering and T.Gladmen; (1967), J.Iron Steel Inst.
17. .P.Baldinger, G.Posh and A.Kneissl, (1994), Practical Metallography, vol 31.
18. .D.Y. Choi, (1992), M.Phil Thesis University of Sheffield
19. L.J.Cuddy and J.C.Raley, (1989). Met Trans A, Vol 14A
20. J.H.Beynon and C.M.Sellars, (1985), American Society for Testing and Materials 1916 Race Street, Philadelphia, PA. 19103.

The Properties and Kinetics of Carburized Duplex Stainless Steel

Nik Rozlin Nik Masdek¹, Iswadi Jauhari^{1*}

¹ Department of Mechanical Engineering,
Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia
Tel: +019-3452554, Fax: +603-7967 5317, E-mail: nikrozlin@yahoo.com

Abstract

In this study, the carburizing process was conducted on duplex stainless steel (DSS). Two types of DSS with different microstructures were carburized – the as-received DSS with coarse microstructure and the fine microstructure DSS. The carburizing process was carried out at temperatures ranging from 1123 K to 1223 K for various durations. The formation of hard carbon layer was confirmed by XRD analysis. Metallographic studies and scanning electron microscope (SEM) revealed a uniform, dense and smooth morphology of carbon layer produced on all of the carburized specimens. In addition, the case depth of carburized layer was examined and measured using metallographic technique and SEM. The carburized layer thickness formed at the carburized surface of carburized as-received DSS was in the range of 15 μm to 40 μm while a much higher layer thickness was produced from the carburized fine grain DSS which is in the range of 16 μm to 60 μm . The surface hardness value for carburized as-received DSS was between 350 HV to 1200 HV while for the carburized fine grain DSS a significant increase in surface hardness was obtained in the range of 420 HV to 1600 HV. The kinetics of this process in terms of carbon diffusion and its variation with processing time and temperature was determined using the Arrhenius equation. The activation energy (Q) for both carburized as-received DSS and carburized fine microstructure DSS was determined as 218 kJ/mol and 198 kJ/mol respectively.

Keywords:

Carburizing, Duplex Stainless Steel, Surface Hardness, Activation Energy

1. Introduction

The carburizing process which is a surface hardening technique for steel have been well developed and widely used in various industrial applications due to its superior combinations of mechanical properties. This process can be carried out whether in a solid, liquid or gaseous medium. Many studies have been carried out in carburizing various steels such as austenitic stainless steel, low carbon steel and pure iron [1-3].

Generally in pack carburizing, the carburizing agent is in powder form. It can be carried out under inert atmosphere where the specimens are embedded into the carburizing powder inside a heat-resistant box. The diffusion of carbon atoms into steel results in formation of iron carbides and the thickness or case depth of the carburized layer is determined by the temperature and time of the carburizing process.

The duplex stainless steel (DSS) has wide applications in industrial equipments such as rotors, fans, shafts and impeller blades. The mixed microstructure of DSS with about equal proportions of ferrite and austenite guarantees higher resistance to pitting and stress corrosion cracking in comparison with conventional stainless steel [4]. Therefore, it will be a great advantage if through the carburizing treatment it can further increase the surface hardness and other mechanical properties such as wear and corrosion resistance which will definitely benefit the surface engineering area.

2. Experimental Procedure

2.1 Substrate Material

Duplex stainless steel (JIS SUS329J1) was used as the substrate material. The chemical composition of the material

is presented in Table 1. In this study, DSS with two different grain sizes were analyzed which is the as-received DSS with coarse microstructure as shown in Figure 1 with the average grain size of 30 μm and the fine microstructure DSS as shown in Figure 2 with the average grain size of 5 μm . To obtain the fine grain microstructure, the as-received DSS was initially solution-treated at 1573 K for 1 hour followed by water quenching. The solution treated DSS was then cold-rolled to a plate through a reduction area of 75%. Specimens for both of the processes were cut from the thermo-mechanically treated plate to a dimension of 10 x 10 x 8 mm.

Table 1. Chemical Composition of Duplex Stainless Steel (JIS SUS329J1) in wt %

C	0.06
Si	0.42
Mn	0.30
P	0.03
S	0.06
Ni	4.18
Cr	24.5
Mo	0.49
Fe	Balance

Figure 1 shows the coarse DSS with the average grain size of 30 μm

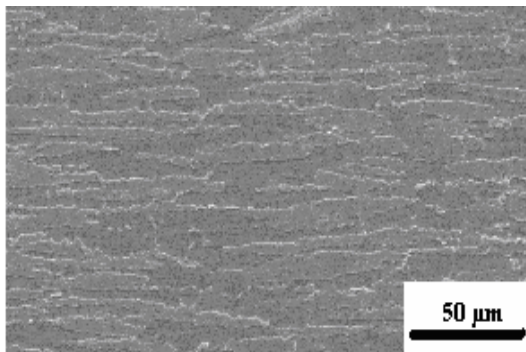


Figure 1. Coarse DSS with the average grain size of 30 μm

Figure 2 shows the fine microstructure DSS with the average grain size of 5 μm

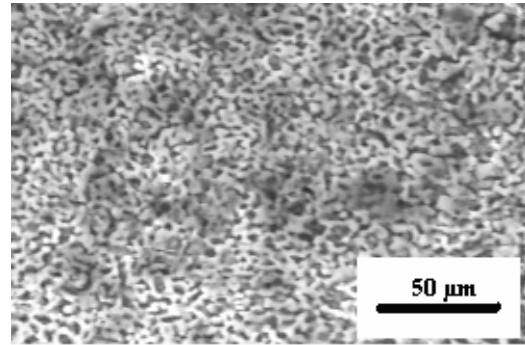


Figure 2. Fine microstructure DSS with the average grain size of 5 μm

2.2 Carburizing Process

Prior to the carburizing process, all specimens were ground by emery paper and cleaned by alcohol in order to remove oxide layers and irregularities at the outer surface in order to enhance carbon uniformity [5]. The specimen was packed in a sealed stainless steel container filled with Wilcarbo powder (carburizing agent) as shown in Figure 3. The carburizing process was then conducted in a furnace with controlled atmospheric condition for 2 to 8 hours at temperatures from 1123 K to 1223 K. After the carburizing process, the specimens were air-cooled to room temperature. Figure 4 shows the process diagram for both carburizing processes.

Figure 3 shows the schematic diagram of carburizing container containing the specimen to be carburized

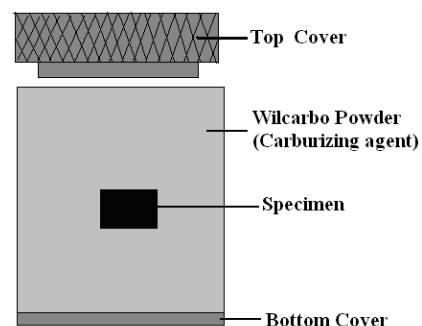


Figure 3. Schematic diagram of carburizing container containing the specimen to be carburized

Figure 4 shows the process diagram of the carburizing process

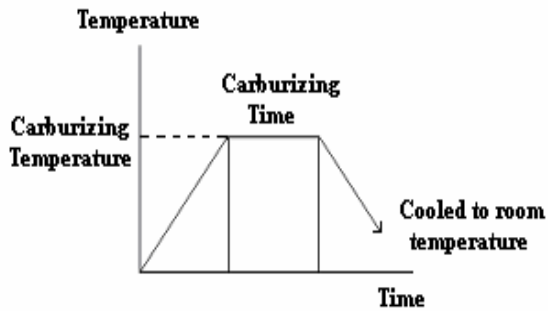


Figure 4. Process diagram of the carburizing process

2.3 Carbon Layer Thickness and Microstructure Characterization

The carburized specimen was ground with emery paper up to 1200 grit and polished until a diamond finish was obtained. The microstructure of carburized DSS was then revealed by using a special echant containing hydrochloric acid (HCl) saturated with ferric chloride (FeCl_3).

The morphology and thickness of carbon layer formed on the surface of carburized DSS was observed by optical microscope and scanning electron microscope (SEM). X-ray diffraction (XRD) analysis was used to determine the presence of carbide phases on the carburized surfaces.

2.4 Hardness Measurement

The surface hardness of carburized specimens was measured using an Instron microhardness tester fitted with a Vickers diamond indenter with a load of 2N. Microhardness was measured at various locations on the carburized surface and also the cross-sectional part in order to measure the hardness variation from the hard carbon layer to the core of the carburized specimen.

3.0 RESULTS AND DISCUSSIONS

3.1 X-ray Diffraction

The presence of carbides on the surface of the carburized specimens was determined by using a Philips X'Pert MPD PW3040 XRD with $\text{CuK}\alpha$ radiation at 1.54056 Å X-ray wavelength. The specimens were scanned from 10 to 80° 2 θ angle at a step size of 0.020 and a count time of 1.5 s at each step. Figure 5 demonstrates the X-ray diffractions of the DSS before and after carburizing process. From the relative peak intensity in the XRD pattern, the presence of phases FeC, Fe_2C and Fe_3C was detected.

3.2 Microstructure and Carbon Layer Thickness

Optical and SEM examinations of both carburized as-received DSS and fine microstructure DSS revealed that similar morphologies of carbon layer were obtained. A uniform, smooth and dense coating was formed at the surfaces of all carburized DSS. Figure 6 demonstrates the cross sectional view of the uniform, smooth and dense morphology of the carburized layer formed at the fine microstructure DSS surface.

Table 2 tabulates the carbide layer thickness of DSS with two different grain sizes at different time and temperature. It can be seen that at 1123K the carbide layer thickness for both as-received DSS and fine grain DSS are more or less equal to each other which ranges from 15 μm to 30 μm . However, the carbide layer thickness is seen to increase with increased temperature and time. This indicated that the carbide layer thickness depends strongly on the carburizing time and process temperature. It can be also seen from the table that the highest carbide layer thickness is achieved by the fine grain DSS. This is due to the fine grain microstructure of DSS which enhances the movements of carbon atoms into the material as it provides a larger number of grain boundary diffusion paths.

Figure 5 shows the X-ray diffraction pattern of fine microstructure DSS before and after carburizing

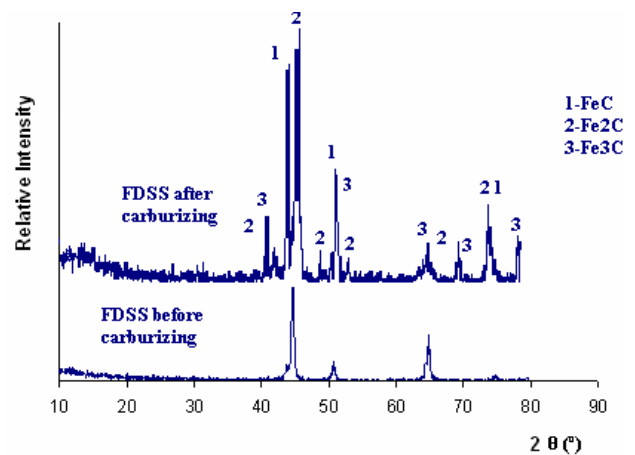


Figure 5 X-ray diffraction pattern of fine microstructure DSS before and after carburizing

Figure 6 shows a cross-sectional view of the carburized layer formed at the fine microstructure DSS surface

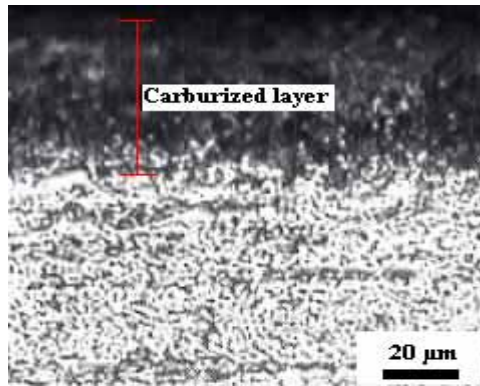


Figure 6 Cross sectional view of the carburized layer formed at the fine microstructure DSS surface

Table 2 Carbide layer thickness

Temperature [K]	Time [hours]	Fine grain DSS	As-received DSS
1123	2	16.1	15.5
	4	23.6	20.8
	6	25.5	24.6
	8	29.1	28.7
1173	2	19.8	18.9
	4	25.6	23.6
	6	35.5	28.5
1223	2	30.2	21.1
	4	43.1	27.1
	6	53.2	30.2
	8	65.0	40.1

3.3 Surface Hardness

The initial hardness value of the as-received DSS and fine microstructure DSS before carburizing process was 346 HV and 426 HV respectively. After the carburizing process, the surface hardness of the as-received DSS increased up to 1223 HV as shown in Figure 7. A much higher surface hardness value was obtained for the fine microstructure DSS which reached 1662 HV after the carburizing process as indicated in Figure 8.

Figure 7 shows the surface hardness of carburized coarse DSS

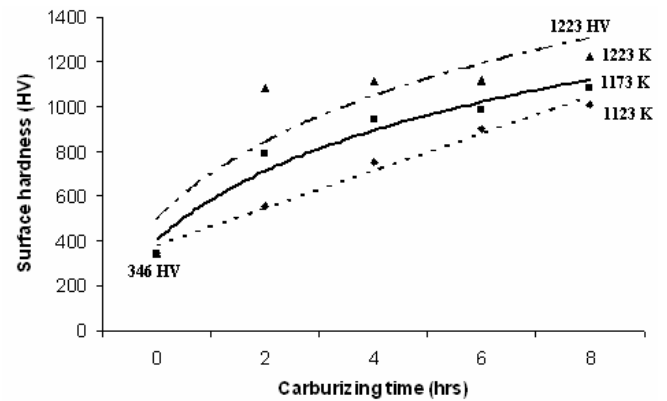


Figure 7 Surface hardness of carburized coarse DSS

Figure 8 shows the surface hardness of carburized fine microstructure DSS

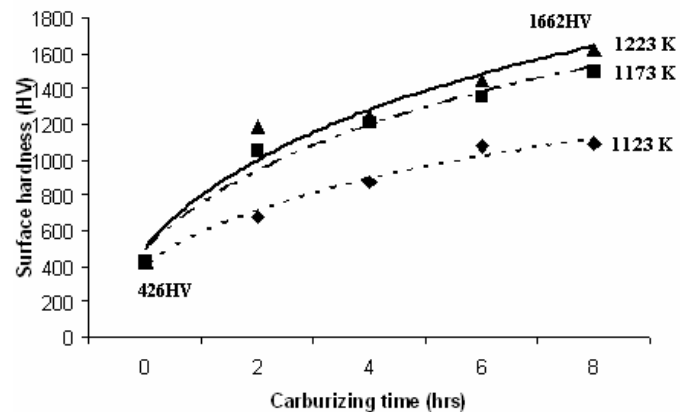


Figure 8 Surface hardness of carburized fine microstructure DSS

3.4 Process Kinetics

It is known that the thickness of carbon layer is mainly dependant on the diffusion of carbon atoms into the substrate material. Therefore, the activation energy of carburizing process can be determined using the thickness of carburized layer formed.

The thickness of carbon layer varies with time and follows the parabolic law as below:

$$d^2 = Kt$$

(1)

where d is the carbon layer thickness (μm), t is carburizing time (s) and K is carbon growth rate constant or diffusion coefficient of carbon into the matrix.

Figure 9 shows the square of carbide layer thickness versus carburizing time of fine microstructure DSS

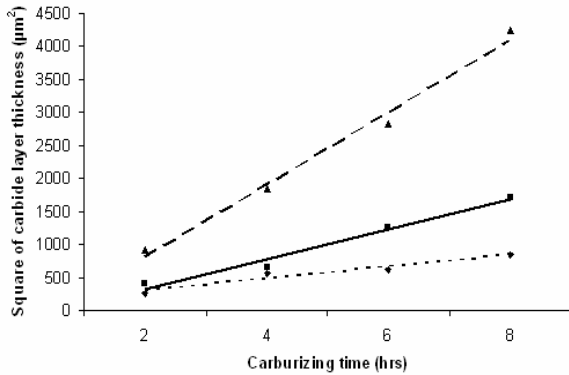


Figure 9 Square of carbide layer thickness versus carburizing time of fine microstructure DSS

Figure 10 shows the square of carbide layer thickness versus carburizing time of as-received DSS

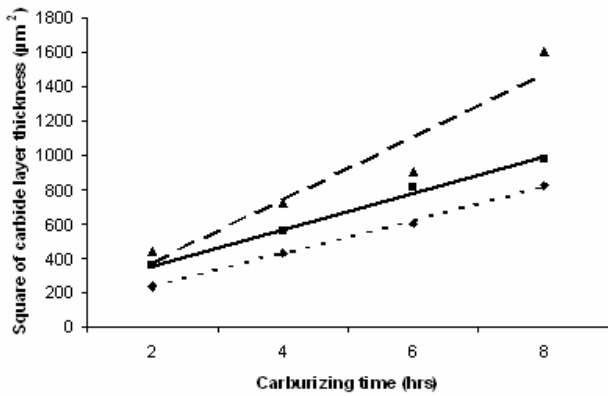


Figure 10 Square of carbide layer thickness versus carburizing time of as-received DSS

Figure 9 and Figure 10 indicates that the square of carbon layer thickness for both fine microstructure DSS and as-received DSS changes linearly with time. Therefore, the relationship between the diffusion coefficients, K ($m^2 s^{-1}$), activation energy, Q ($J mol^{-1}$) and carburizing temperature, T (K) can be expressed by an Arrhenius equation as follows:

$$K = K_o e^{\frac{Q}{RT}} \tag{2}$$

where K_o is the pre-exponential constant and R is the gas constant ($8.314 J mol^{-1} K^{-1}$). Taking the natural logarithm of Equation (2), Equation (3) can be derived as follows:

$$\ln K = \ln K_o + \left(\frac{Q}{R} \right) \left(T^{-1} \right) \tag{3}$$

As shown in Figure 11 and 12, the plot of natural logarithm of carbon growth rate ($\ln K$) versus reciprocal carburizing temperature (T^{-1}) is linear. The slope of the straight line will determine the activation energy (Q) of the carburizing process.

Figure 11 shows the natural logarithm of carbon growth rate ($\ln K$) versus Reciprocal carburizing temperature (T^{-1}) for carburized as-received DSS

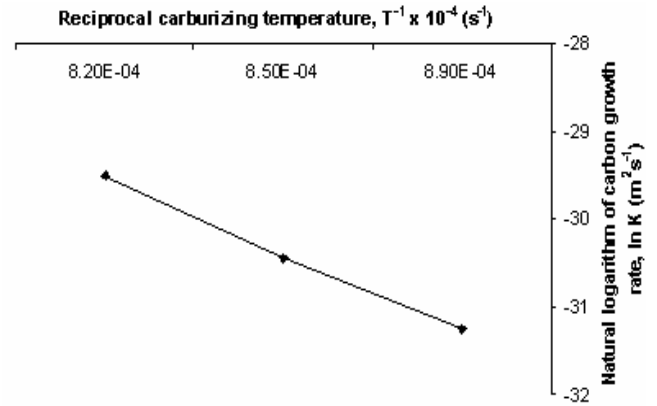


Figure 11 Natural logarithm of carbon growth rate ($\ln K$) versus Reciprocal carburizing temperature (T^{-1}) for carburized as-received DSS

Figure 12 shows the natural logarithm of carbon growth rate ($\ln K$) versus Reciprocal carburizing temperature (T^{-1}) for carburized fine microstructure DSS

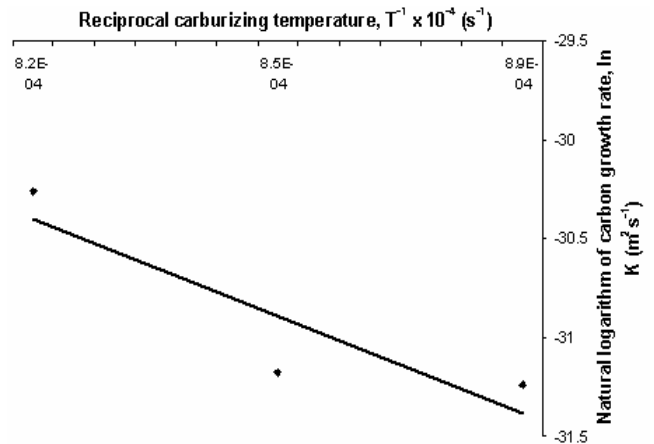


Figure 12 Natural logarithm of carbon growth rate ($\ln K$) versus Reciprocal carburizing temperature (T^{-1}) for carburized fine microstructure DSS

It was reported that the value of activation energy for conventional carburizing process was 205.57 kJ/mol [6]. The activation energy for both the carburized as-received DSS and fine microstructure DSS was determined as 218.87kJ/mol and 198.58 kJ/mol respectively. An activation energy decrease of about 47 kJ/mol was achieved through carburizing of fine microstructure DSS. This indicates that the fine grain size of DSS was able to reduce the energy required for the carbon atoms to diffuse into the substrate material, thus reducing the time and cost while enhancing the surface properties of DSS as well.

4.0 CONCLUSIONS

In the present study, carburizing was successfully conducted on the fine microstructure DSS and also the as-received DSS with coarse microstructure. From the results obtained, this study can be concluded as follows:

1. Both the carburized fine microstructure DSS and as-received DSS revealed a thick, uniform and dense morphology of carburized layer obtained on all of the carburized DSS.
2. The fine microstructure of DSS accelerates the diffusion of carbon atoms into the substrate material, thus, resulting in a much thicker carburized layer.
3. The phases formed on the carburized fine microstructure DSS was a mixture of FeC, Fe₂C and Fe₃C as confirmed by XRD analysis.
4. A remarkable increase in surface hardness was obtained through carburizing of the fine microstructure DSS compared to the surface hardness of carburized as-received DSS.
5. The activation energy of the carburizing process of fine microstructure DSS and as-received DSS was determined as 198.58 kJ/mol and 218.87 kJ/mol respectively.

Acknowledgements

The authors would like to thank Vote F PJP no F0134/2005A for funding the research.

References

- [1] B.S. Suh and W.J. Lee: Thin Solid Films 295 (1997) p. 185-192
- [2] P. Kula, R. Pietrasik and K. Dybowski: Vacuum carburizing-process optimization, *Materials Processing Technology* 164-165 (2005) 876-881
- [3] M. Tarakci, K. Korkmaz, Y. Gencer and M. Usta: Plasma electrolytic surface carburizing and hardening of pure iron. *Surface and Coatings Technology*, 199 (2005) 205-212
- [4] Cabrera, J.M., Mateo, A., Llanes, L., Prado, J.M., Anglada, M.: Hot deformation of duplex stainless steels. *Journal of Materials Processing Technology*. 143-144 (2003) 321-325
- [5] Sun, Z., Kuo, M., Annergren, I and Pan, D., (2003), "Effect of dual torch technique on duplex stainless steel welds", *Material Science & Engineering A*, 356, pp. 274-282.
- [6] S.S. Iskander and E.B. Saleh: *Carburizing of steels*. Surface Technology. Volume 12, 4 (1981) 341-349

The Pretreatments and Microwave Power Effects on Nucleation and Growth of Polycrystalline Diamond Coated Si₃N₄

E. Hamzah.¹ A Purniawan.^{1*}, M. R. M. Toff².

¹ Department of Materials Engineering, Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Malaysia
Telp: +60-7-5534563, Fax: +60-7-5576820 E-mail: esah@fkm.utm.my

² Nanomaterials Programme
Advanced Materials Research Center (AMREC), Lot 34, Jalan Hi-Tech 2/3,
Kulim Hi-Tech Park, 09000 Kulim, Kedah, Malaysia.
Telp: +60-4-401715, Fax: +60-4-4033224, Email: mradzit@sirim.my

Abstract

Diamond is the hardest material and high wear resistant considered as thin film was deposited onto surface substrate to improve cutting tools properties. Silicon Nitride (Si₃N₄) cutting tools material base ceramic was used as substrate material and the deposition of diamond coating by using microwave plasma assisted chemical vapor depositions (MPACVD). Two pretreatment processes namely scratching and seeding were conducted. Furthermore, deposition process was conducted to deposit diamond using microwave power 2.75, 3.25 and 3.50 kW on the substrate. The microstructure and morphology of diamond coated Si₃N₄ cutting tools were investigated by Scanning Electron Microscopy (SEM) and Atomic Force Microscopy (AFM). The existence of diamond was verified using X-ray Diffraction (XRD) method. It is found that the seeding effect on diamond nucleation rate is more dense as compared to scratching. Increasing of microwave power has significant effect on the faceting and growing of polycrystalline diamond. It showed that X-ray diffraction analysis is one of the significant tools to verify the presence of diamond on the substrate surface

Keywords: polycrystalline diamond, Silicon Nitride, MPACVD, seeding, scratching

1. Introduction

Diamond is the hardest material and extreme wear resistance, chemical inertness, and thermal conductivity. The excellent properties of diamond, especially polycrystalline diamond are widely used as thin film coating material in cutting tools and protective coating. Microwave plasma assisted chemical vapor deposition (MPACVD) has attracted considerable interest in cutting tools coating [1-3].

The properties of nitride ceramics depend on several factors. Because mainly of the strongly covalent character of this material, it distinguishes itself from other materials by a combination of several advantageous properties, such as high hardness and mechanical strength, a good thermal conductivity and, hence, a high resistance to thermal shocks, and a low liability to oxidation, which in turn, ensures a long service life of the tool edge. It is essential that this material retains all these properties when heated to high temperature. The mechanical strength of nitride ceramics also depends on the size of the pores and fissures present [5]. At the present time, widely used substrate materials for CVD diamond coating are silicon (Si), silicon carbide (SiC), silicon nitride (Si₃N₄) and tungsten carbide (WC-Co). As polycrystal silicon is easy to cut and grind, it is the best substrate material for CVD diamond coating [4]. K. Inderjeet et al. [7] reported that the deposition of

diamond crystallites was less homogenous in both the Si wafer and WC if compared to the Si₃N₄.

The diamond deposition on non diamond substrate surface dependent on several factors such as pretreatment, substrate temperature, chamber pressure, gas mixture, and power. The mechanical treatment using microscratching (seeding with diamond particles) has an effect on adhesion and quality of diamond coating [6]. K. Mallika et al [1] reported grain size and morphology of the diamond coatings were determined primarily by the % CH₄ concentration in the CH₄/H₂ gas mixture. Within the deposition temperature range used in this investigation, substrate temperature did not appear to have any significant influence on the grain size and morphology of the diamond. However, it had notable effects on the quality of diamond coating.

The objective of the study is to investigate influence of mechanical pretreatment namely scratching and seeding and microwave power on microstructure and morphology of polycrystalline diamond coating on Si₃N₄ cutting tools material. The microstructure and morphology was analyzed by using SEM and AFM. The existence of diamond was investigated using XRD analysis.

* Corresponding author E-mail: agung_purniawan@yahoo.com

2. Experimental Detail

Silicon Nitride (Si_3N_4) was used as substrate material. The composition and main crystalline phases as compound on the uncoated Si_3N_4 substrate was determined by using XRD technique. Samples were cut to dimension 12 mm x 12 mm x 3 mm. Scratching process by using polishing machine and diamond powder grit 1 μm was carried out for 10 min followed by ultrasonically cleaned in acetone for 15 min to remove oil and grease contaminations on the substrate prior to deposition. The other treatment method is seeding by ultrasonic process and diamond powder grit 15 μm for 60 min.

The diamond films were grown on Si_3N_4 substrates by the microwave plasma assisted chemical vapor deposition (MPACVD) technique in Penta Vacuum (2.45 GHz, 6 kW) system under the following conditions is shown in Table 1. The layout of the MPACVD equipment is shown in Figure 1.

Table 1. Experimental conditions for diamond deposition

Parameter	Value
Microwave power	2.72, 3.25 and 3.50 KW
Pressure	20 - 50 torr
Total gas flow rate	~ 198 – 204 sccm
Gas mixture	1.5 – 2% CH_4 in H_2
Total deposition time	7 hours

The deposition process in the experiment was conducted under three step processes. The first step is plasma etching process using H_2 in the chamber. The second step is nucleation growth process. The third step polycrystalline diamond growth process. The finally process is cooling process.

SEM and AFM were used to investigate the microstructure and morphology of diamond coating and existence and form of diamond was investigated using XRD.

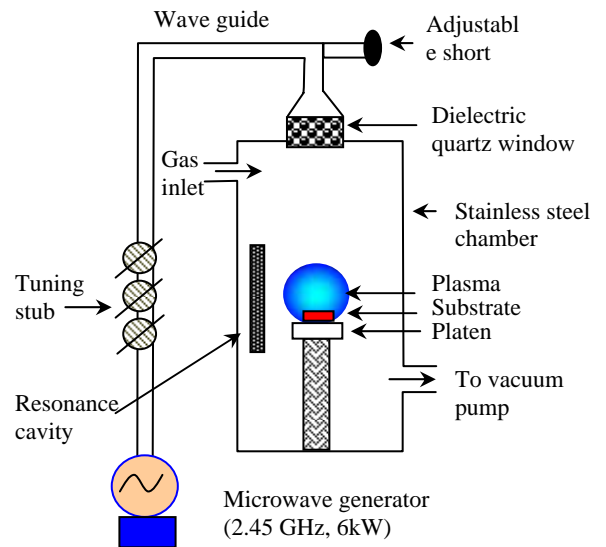


Figure 1. Layout of the MPACVD system

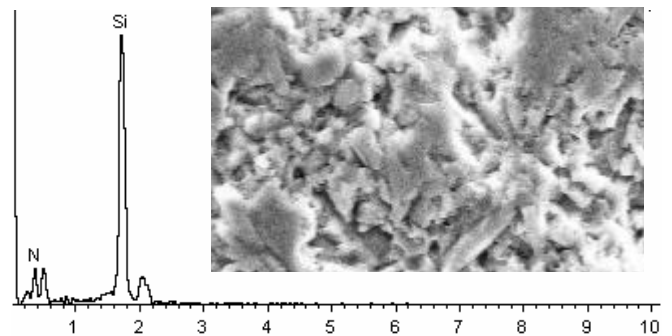
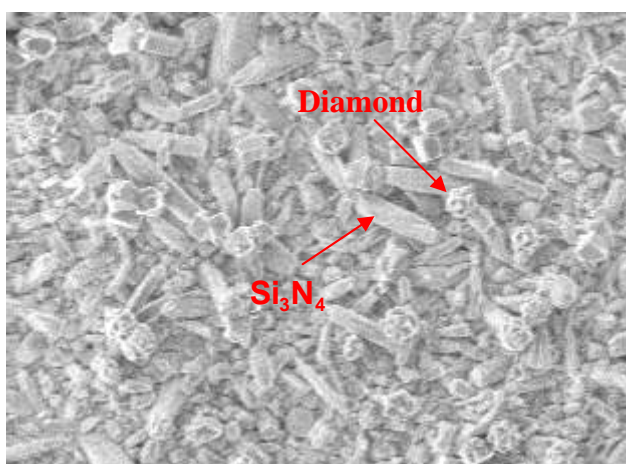
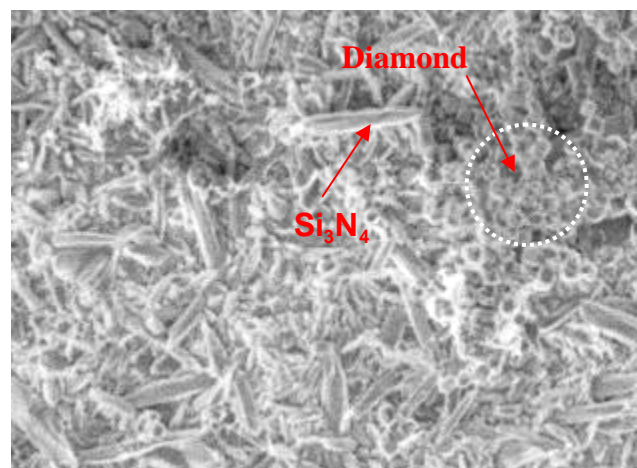


Figure 2. The EDX spectrum and SEM micrograph of uncoated Si_3N_4



(a)



(b)

Figure 3. SEM micrograph of coated Si_3N_4 with pretreatment (a) scratching (b) seeding

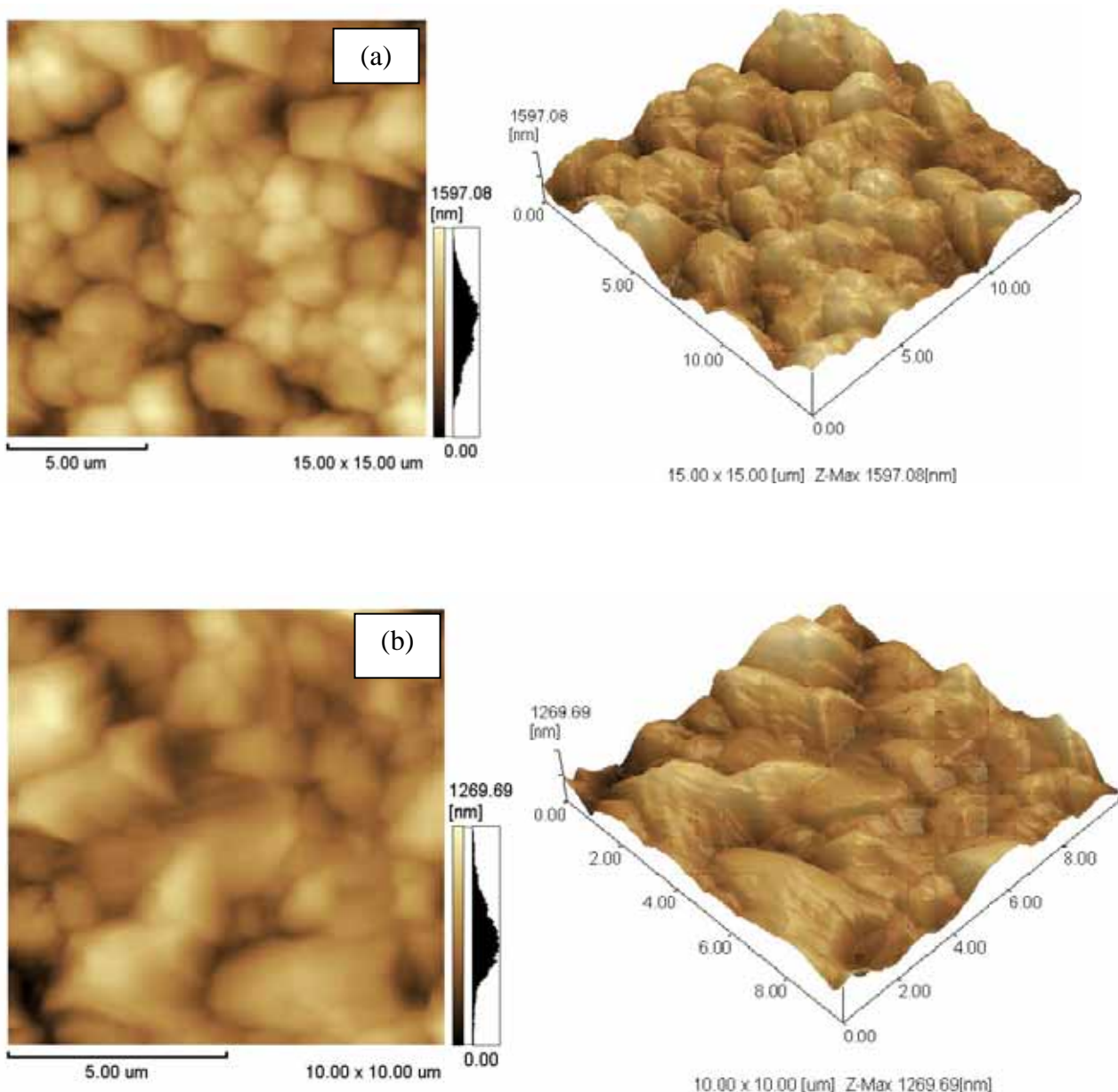


Figure 4. The topography of diamond coated Si_3N_4 using AFM (a) 2.75 kW microwave power (b) 3.25 kW microwave power.

3. Test Results and Discussion

The Si_3N_4 uncoated composition is shown in Figure 2. Figure 3 shows the scanning electron microscopy (SEM) micrograph of the diamond surface after different pretreatments. Seeding with diamond powder 15 μm ultrasonically for 60 min shows more significant effect than scratching by using polishing machine with diamond powder 1 μm for 10 min shows the effect in diamond nucleation of diamond nucleation (Fig. 3a and Fig. 3b). The microwave power was found to have minimum influence [1]. However, in the present study the microwave power is

a significant parameter. The morphology of diamond coated Si_3N_4 is investigated using atomic force microscopy (Fig. 4). An increase of microwave power can lead to an increase morphology shape and also facets of the crystals.

In the existence of diamond was investigated using X-ray diffraction. Figure 5.(a) shows the XRD peak of uncoated Si_3N_4 . The existence of diamond is shown in Fig.5.(b), the pattern of diamond phase indicated clearly the existence of diamond on the coated Si_3N_4 by scratching pretreatment process.

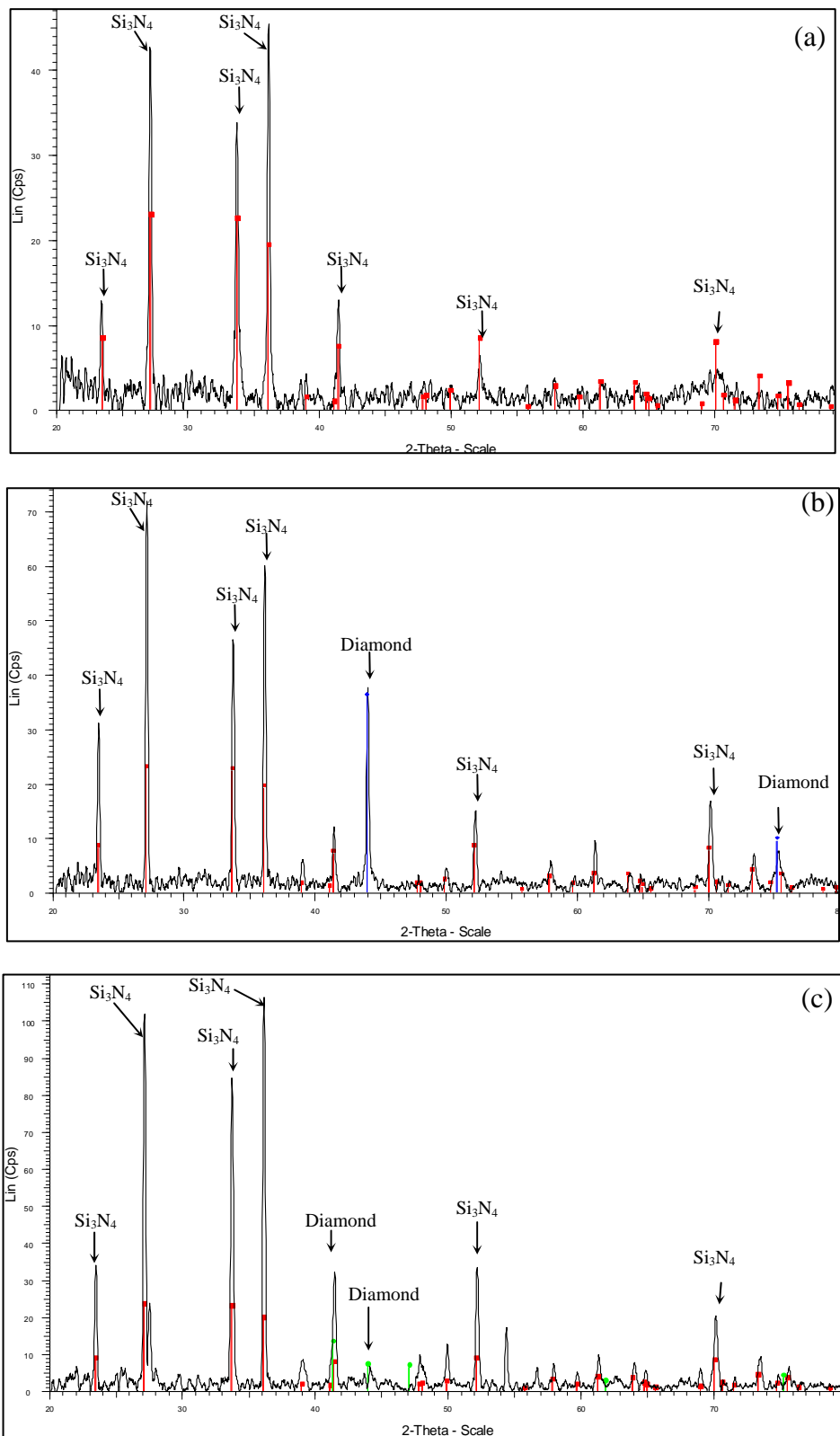


Figure 5. X-ray diffraction pattern of Si_3N_4 (a) uncoated (b) coated on scratched sample (c) coated on seeded sample.

4. Conclusion

The presence of diamond on surface substrate was verified by using XRD technique. The seeding treatment is found to have more dense effect on diamond nucleation process than scratching. The microwave power has an effect on faceting and growing of crystal.

Acknowledgments

The work supported by the Ministry of Science, Technology and Innovation of Malaysia under the Intensification of Research in Priority Areas (IRPA) is gratefully acknowledged. [03 – 02 – 14 – 0001 PR 0074/ 03 – 02]

References

- [1] K. Mallika, R. Komanduri. 2001. Low pressure microwave plasma assisted chemical vapor deposition (MPCVD) of diamond coatings on silicon nitride cutting tools. *Thin Solid Films* 396: 145–165
- [2] A.J.S. Fernandes, E. Salgueiredo, F.J. Oliveira, R. Silva, F.M. Costa. 2005. Directly MPCVD diamond-coated Si_3N_4 disks for dental applications. *Diamond and Related Materials* 14: 626– 630
- [3] Jih-Kun Yan, Li Chang. 2005. Microwave plasma chemical vapor deposition of cone-like structure of diamond/ SiC/Si on Si (100). *Diamond & Related Materials* 14: 1770 – 1775
- [4] T.Abe, T.T Akagi , Z.M.Sun , T.Uchimoto , J.Makino , H.Hashimoto. 2004. Machinable ceramic substrate for CVD diamond coating. *Diamond and Related Materials* 13: 819–822
- [5] A.Olszyna, J.Smolik. 2004. Nanocrystalline diamond-like carbon coatings produced on the Si_3N_4 –TiC composites intended for the edges of cutting tools, *Thin Solid Films* 459: 224–227
- [6] F. Deuerler, H. van den Berg, R. Tabersky, A. Freundlieb, M. Pies, V.Buck. 1996. Pretreatment of substrate surface for improved adhesion of diamond films on hard metal cutting tools. *Diamond and Related Materials* 5: 1478 – 1489
- [7] K. Inderjeet, S. Ramesh, C.K. Chakrabarty. 1999. Nucleation and growth of polycrystalline diamond particles on ceramic substrates by microwave plasma CVD. *Surface Modification Technologies (SMT XIII), ASM International* Vol. 13 pp. 83 – 90.

An experimental investigation of coated ceramic cutting tools when hard turning cold work tool steel

M. A. Kamely^{1*}, M. Y. Noordin¹, V. C. Venkatesh², D. Kurniawan¹

¹ Faculty of Mechanical Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5534697, Fax: +60-7-5566159, E-mail: kamelysita@yahoo.com

² Faculty of Engineering & Technology
Multimedia University, Jalan Ayer Keroh Lama, 75450 Melaka, Malaysia
Tel: +60-6-2523319, Fax: +60-6-2316552, E-mail: venkatesh@mmu.edu.my

Abstract

Studies have shown that by using the right combination of insert tool geometry and cutting parameters, hard turning can produce the same and even better surface finish compared to the grinding process for finish machining. The implementation of hard turning was made possible by the development of new cutting tools and improvements in machine tool rigidity. Ceramic tools are widely used today in the manufacturing industry for cutting various hard materials, including those for moulds and dies. In the mould and die industries, cold work tool steels are extensively used. The life of a mould or die is determined mainly by its resistance to wear processes, which can be of different nature depending on the application. Therefore, moulds and dies are usually hardened and have excellent surface finish. Newly developed ceramic turning grades are intended to permit versatile use in roughing and finishing applications for a broad spectrum of hard turning. Thus, it becomes very important to quantify the performance of ceramic cutting tools when hard turning cold work tool steel based on tool wear. Tool wear mechanisms were intensively studied by using optical microscope and field emission scanning electron microscope (FESEM). Results indicated that the tool life decreased considerably with the increase in cutting speed and it seemed unlikely that abrasion was the major wear mechanism.

Keywords: Ceramic cutting tool, hard turning, cold work tool steel, wear mechanisms

1. Introduction

The machining of hardened steels using ceramic cutting tools has been widely investigated by researchers today. There are a large number of steel components that usually need a finishing operation after the heat treatment process. For certain applications, hard turning can deliver tolerances and surface finishes better than grinding process. Generally, mould and die making involves an Electron Discharge Machining (EDM) process and a machining operation such as turning and milling. In manufacturing process, mold and die work usually refers to parts produced that are asymmetrical. Mold and die components need to have high dimensional accuracy and excellent surface finish. For extremely tight-tolerance jobs, hard turning can also serve as an effective pre-finishing operation. The requirements for cutting hard materials are good thermal conductivity and hardness of the tool, a large horsepower of the machine tool and good stiffness [1]. Hard machining has become possible by using the range of advanced cutting materials such as PCBN, sialon, mixed ceramics and whisker-reinforced alu-

mina [2]. The use of ceramic cutting tools with appropriate machining parameters on hard turning of cold work tool steel permit a surface roughness ($R_a < 0.8 \mu\text{m}$) corresponding a high dimensional precision without necessity of cylindrical grinding operations [3].

Ceramic materials retain their hardness and stiffness at temperatures above 1000°C . Some of them also exhibit chemical inertness, depending on their high free energy of formation. They are expected to resist abrasive and chemical wear at high temperatures generated during the turning operations [4]. Unlike most metals, hardness levels in ceramics generally remain high at elevated temperature [5]. Among the earliest reports of the machining of hardened steels with geometrically defined cutting tools is that of King and Wheildon [6]. Results showed that hard turning with mixed alumina tools produced specific surface profiles and microstructures, although the R_a roughness values of about $0.25 \mu\text{m}$ can be comparable to those produced by finish grinding [7]. But, besides the abovementioned positive properties, ceramic materials also exhibit low toughness and low thermal conductivity. Therefore, they are prone to brittle fractures induced by mechanical and or thermal stresses. As a consequence, the suitability of ceramic tools for cutting cold work tool steel must be evaluated. Grzesik et al. [8]

* Corresponding Author. E-mail: kamelysita@yahoo.com, Tel: +60-7-5534697, Fax: +60-7-5566159

investigated the contribution to surface roughness generation when turning low chromium alloy steel, heat treated to a hardness of 60 HRC, with ceramic cutting tools and cylindrical grinding. Tool nose radius effects in finish turning of hardened AISI 52100 steels have been investigated [9]. In previous works, the behaviors of different commercial ceramic inserts when cutting carbon steel was investigated and compared with that of traditional carbide based tool [10].

Failure of ceramic tools is generally caused by fracture, especially in an intermittent cutting process, because of their relatively low fracture toughness, tensile strength, and thermal shock resistance. Even in continuous turning process, tool failure by wear is usually accompanied by chipping and flaking in rake face and nose [11]. The wear performances of several types of cutting tool including ceramic have been studied. The results showed that the tool wear mechanism differed in various matching of materials between the cutting tool and the workpiece.

The development of new cutting tools and improvements in machine tool rigidity has provided an alternative for finish machining. Newly developed turning grades are intended to permit versatile use in roughing and finishing applications for broader spectrum of workpiece materials. The new improvements and developments in coating technology have also advanced by supplying new and more wear resistant tool materials.

The use of coated cutting tools to machine various hard to cut materials now represents the state-of-the-art technology. Coatings of cutting tools provide a combination of wear resistance and chemical stability at high temperature to meet the demands of the applications. Cutting tools applied for hard turning such as PCBN are relatively expensive. Ceramic cutting tools are the less expensive choice and their properties may be increased by coating. So, it is important to investigate coated ceramic cutting tools' performance to assure the economic justification for hard turning.

In this study, the performance of coated ceramic cutting tool when turning hardened AISI D2 cold work tool steel (60 HRC) was evaluated based on tool life and wear mechanisms. Coated ceramics remain the 'enabling technology' for hard turning. However, the mechanisms through which they wear are not adequately understood yet.

2. Experimental Details

2.1. Cutting Conditions

Commercially available AISI D2, a type of cold work tool steels, bars of 90 mm diameter and 200 mm length, hardened to 60 HRC were used as workpiece material. The typical composition of this special alloy steel is 1.55 % C, 0.4 % Mn, 11.6 % Cr, 0.8 % Mo, 0.9 % V and 0.3% Si. The hardness of the workpiece was also measured using a Rockwell hardness tester for indicating the homogeneity of the workpiece. The machining operations were carried out on a 5.5 KW MAHO Graziano GR200 CNC slant bed lathe machine with a 432T control unit under dry cutting condition. Before conducting

the machining tests, a thin layer of 0.5 mm was removed from the workpiece surface using a tool with new cutting edge to ensure consistency. The machining tests were carried out at three different speeds while the feed rate and the depth of cut were held constant at a reasonable value based on feedback from the manufacturers. The cutting tool materials used in this study are the commercially available PVD TiN coated aluminum oxide and titanium carbonitride ceramic ($\text{Al}_2\text{O}_3/\text{TiCN}$). The specifications of the cutting tool materials, cutting conditions, and tool holders are shown in Table 1.

Table 1. The specifications of the cutting tools, cutting conditions, and tool holders used in this study

Tool holder	:	MCLNL 1616H 12
Insert	:	CNGA120408T01020
Nose radius	:	0.8 mm
Chamfer angle	:	20 ⁰
End cutting edge angle	:	5 ⁰
Side rake angle	:	-5 ⁰
Side cutting edge angle	:	5 ⁰
Rake angle	:	-5 ⁰
Inclination angle	:	5 ⁰
Cutting speed	:	100, 140, 200 m/min
Feed rate	:	0.06 mm/rev
Depth of cut	:	0.4 mm
Coolant	:	none (dry)

The flank wear was measured using tool maker's microscope. The surface roughness was measured using a portable surface roughness tester, Rank Taylor Hobson Surtronic 3+. A cut-off length of 0.8 mm was used. The centre line average, R_a , was taken to represent the particular test. The surface roughness tester was periodically calibrated using a standard reference specimen. Three measurements were taken for each length to obtain the average surface roughness value. Tool life was determined by measuring maximum flank wear width, VB_{max} , of the tool. The cutting was stopped when the tool flank wear, VB_{max} , reached 300 μm . This criterion is used to achieve dimensional accuracy and surface roughness required for finish turning. The tool's surfaces were further examined under different magnifications using optical microscope and field emission scanning electron microscope (FESEM) in order to get detail results of the wear mechanisms.

2.2. Tool wear investigation

Tool wear was measured at the flank face of the inserts without dismounting it from the tool holder. Tool wear was measured by placing the tool holder underneath a Nikon digital toolmaker's microscope with magnification of 10x (Model CM-6F). The special jig was fabricated in order to stand the tool holder and make sure the flank face is at the same angle every time measurement was taken. The perpendicular distance between the respective positions on the wear profile to the original cutting edge were measured by moving the x-y table. The distance measured was read from a digital read out provided at the microscope and then recorded. Upon measuring, the worn insert was also

observed using a stereo microscope (Zeiss, Type Stemi 2000-C). The worn insert was observed from various positions in order to view the wear occurring on the tool's major and minor flanks as well as on the rake face. The observed images were captured by using the imaging software.

3. Results and discussion

3.1. Mechanism of tool wear

Cutting tools may fail by several factors, such as brittle fracturing, plastic deformation, or gradual wear. The tool wear depends on the tool, workpiece material (physical, mechanical, and chemical properties), tool geometry, cutting parameters, cutting fluids, etc. [12]. In finish machining, the tool wear zone only occurs at the corner of the tool nose radius.

3.1.1. Flank wear

A typical flank wear grooves were observed in this present investigation (Fig. 1). These grooves are believed as the result from abrasion by hard carbide particles contained in the workpiece. Hard particles of work piece microstructures have a high influence on the tool wear process. These pheno-

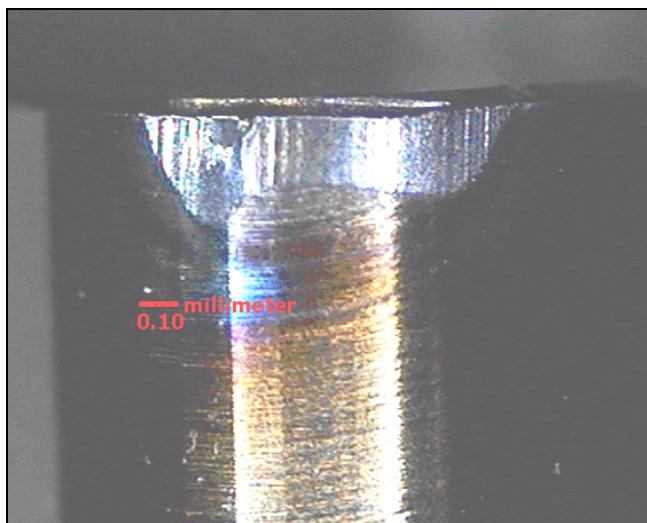


Fig.1. A wear pattern on worn cutting tools at the end of tool life (cutting speed = 140 m/min, feed rate = 0.06 mm/rev, depth of cut = 0.4 mm).

mena are the same as observed by other authors [13]. The tool edge rubbed the newly generated work piece machined surface. The pressures at the sliding interface were exceptionally high. This increased the temperature and compressive stresses near the cutting edge, which would lead to the shortening of the tool life. The edge chipping and removal of substrate material was also observed.

Tool wear progression with cutting time (Fig. 2) shows that, after high initial wear rate, flank wear width increased in a linear behaviour. The results indicated that the flank wear rate seems to be linearly proportional to the cutting speed. As the cutting speed increased from 140 m/min to

200 m/min, flank wear rates show very minor changes. Chipping at the nose radius was observed at all cutting speeds and was more severe at the high speed.

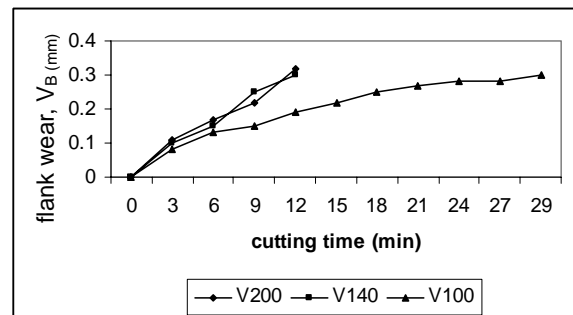


Fig. 2. Tool wear against cutting time

3.1.2. Crater wear

At the crater, it was also found that the built-up layer underwent molten state possibly due to low melting temperature of the newly formed chemical compounds, evidenced by the resolidified built up layer being squeezed out at the direction of workpiece flow (Fig. 3). This formed layer consisted of workpiece materials. Built up layer played an important role in the wear behavior of ceramic tools.

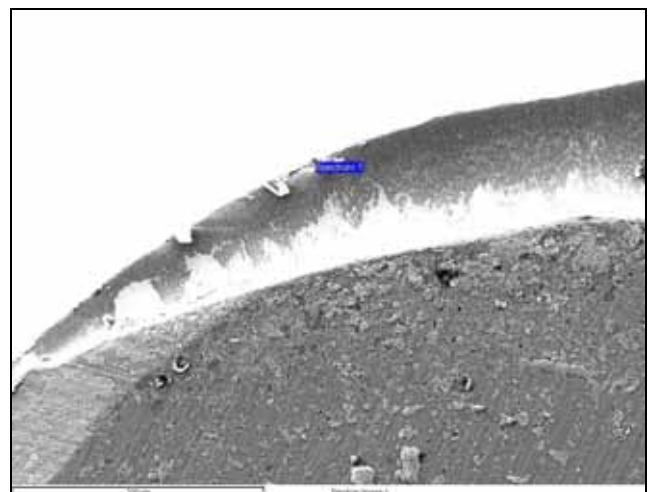


Fig. 3. Adhered layer formed at the crater

It appeared that the wear of the tested ceramic tool was temperature dependent. It was evident that workpiece material deformed and adhered at the crater (Fig. 3), and this may result in reduced friction between the tool and the newly machined surface. Previous study suggested that the coating possibly provide a 'break-away' layer that delayed the exposure and wear of the tool's substrate thus prolong tool life [14]. Once the hard coating worn off and exposed the substrate, the tool-work contact temperature and the cutting forces would rise rapidly, leading to nose deformation and accelerated wear. The TiN coating was worn away rapidly at high speeds via discrete plastic deformation (Fig. 4). Plastic deformation was usually described as superficial or discrete in nature as it is limited to the outermost μm or so of the tool surface [15,16]. When this happened, the crater would

quickly be filled by workpiece material. For most of the tools, the TiN coatings remained intact throughout the entire duration of the experiment.

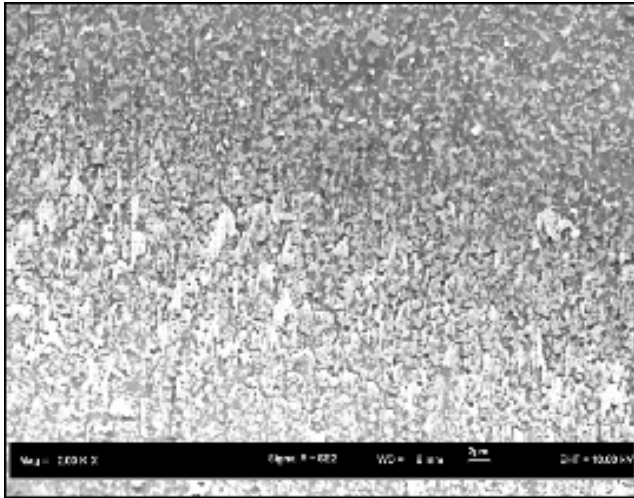


Fig. 4. FESEM view of plastically deformed coating in the direction of workpiece rotation (cutting speed = 140 m/min, feed = 0.06 mm/rev, depth of cut = 0.4 mm).

It can be noticed that at the rake face, the crater developed only at the chamfer area and appeared more severe when the cutting speed was higher. The binder of the tool was abraded by hard carbide particles of AISI D2 cold work tool steel that led to $\text{Al}_2\text{O}_3/\text{TiCN}$ grains detachment from the bond. In addition, an increase in cutting speed led to an increase in cutting temperature, which contributes to this, and thus, the tool-wear was accelerated. The temperature at the crater region can be more than 1000°C up to the melting point of the material being cut and depending on the condition and material properties [17]. The life of ceramic tool insert thereafter became rapidly reduced. An increase in cutting speed led to the shortening of the tool life. The high cutting speed induced high cutting temperatures and the abrasive particles rapidly eliminated the coating layers. Moreover, the chemical wear damaged the binder phase. After the initial rapid wear, all the cutting speeds resulted in similar, reasonably stable wear rates. In addition, an adherent layer appeared on tool crater. All the three tools tested suffered coating loss. Ridge-and-furrow marks at the remaining coating were seen under all the conditions used in the present tests. This might be caused by the softening of the TiN layer, rendering it more susceptible to plastic deformation. In tools tested at high speeds, the coating is worn away rapidly, exposing the substrate beneath. This ridge-and-furrow appearance has previously been reported [18], and the mechanism responsible was termed 'discrete plastic deformation'. Consequently, $\text{Al}_2\text{O}_3/\text{TiCN}$ particles were plucked out due to the loss of binder and further caused accelerated abrasive wear. From the observations, it also showed that the tool wear was associated with localized chipping of the cutting edge. The tools failed by edge chipping which might be caused by the accumulation of the abovementioned factors to the cutting tool. There was only a uniform wear pattern at the flank, but on the rake face, the formed crater was so wide.

4. CONCLUSIONS

Coatings primarily increase wear resistance and may also reduce cutting forces and temperature at the tool edge, thereby indirectly affect the deformation and fracture behaviour of the tool. The performance of coated ceramic cutting tools when machining AISI D2 cold work tool steel (60 HRC), was investigated experimentally in terms of wear mechanisms and the following conclusions can be drawn:

1. A PVD TiN coating over an aluminum oxide and titanium carbonitride ceramic ($\text{Al}_2\text{O}_3/\text{TiCN}$) were suitable for use in machining AISI D2 cold work tool steel of 60 HRC. It is shown that for all cutting speeds, at the selected cutting parameters, the surface finishes obtained fulfilled the finish machining requirement.
2. Tool wear increased gradually for all cutting speeds. The main wear mechanism of the tested tools was abrasion with all the three tools tested suffered coating loss through discrete plastic deformation and attrition.

REFERENCES

1. Tönshoff, H.K., Bussmann, W. and Stanske C. 17-18 Sept. 1986. Requirements on tools and machines when machining hard materials, in; S.A. Davies (Ed.), Proc. 25th MTDR Conf. 349-357. Manchester, Macmillan, London.
2. Tönshoff, H.K., Wobker, H.G. and Brandt, D. 1995. Hard turning-influence on the workpiece properties. Trans. NAMRI/SME 23:215-220.
3. Grzesik, W., Wanat, T. and Brol, S. 2003. A study of roughness profile generation in hard turning and grinding. Sixth international ESAFORM conference on material forming, 579-582. Salerno, Italy.
4. Casto, S.L., Valvo, E.L., Lucchini, E., Maschio, S., Piacentini, M. and Ruisi, V.F. 1999. Ceramic materials wear mechanisms when cutting nickel-based alloys. Wear 227-233.
5. Whitney, E.D. 1994. Ceramic cutting tools-materials, development, and performance, Library of congress catalogue, in publication data 13-215.
6. King, A.G. and Wheildon, W.M. 1966. Ceramics in Machining Processes, Academic Press, New York.
7. Grzesik, W. and Wanat, T. 2005. Comparative assessment of surface roughness produced by hard machining with mixed ceramic tools including 2D and 3D analysis. Journal of Materials Processing Technology 169:364-371.
8. Grzesik, W., Wanat T. and Brol, S. 2003. A study of roughness profile generation in hard turning and grinding. Sixth international ESAFORM conference on material forming, 579-582. Salerno, Italy.
9. Chou, Y. K., and Song, H. 2001. Hard turning with different nose-radius ceramic tools. Technical paper, Society of Manufacturing Engineers.
10. Casto, S.L., Valvo, E.L., Lucchini, E., Maschio, S., Piacentini, M. and Ruisi, V. F. 1995. in: I.M. Low, X.S. Li_Ed., Advanced Ceramic Tools for Machining

- Application: II. Key Engineering Materials, Trans Tech Publications. 114, Chap. 3.:105–134.
11. Xing, A., Jun, Z., Chuanzhen, H. and Jianhua Z. 1998. Development of an advanced ceramic tool material-functionally gradient cutting ceramics, *Materials Science and Engineering A* 248:125–131.
 12. Poulachon, G., Bandyopadhyay, B.P., Jawahir, I.S., Pheulpin, C.S. and Seguin, E. 2003. The influence of the microstructure of hardened tool steel workpiece on the wear of PCBN cutting tools. *International Journal of Machine Tools & Manufacture* 43:139–144.
 13. Ohtani, T. and Yokogawa, H. 1988. The effects of workpiece hardness on tool wear characteristics. *Bulletin of Japan. Society of Precision Engineers*. 22(3): 229-231.
 14. Dawson, T.G. and Kurfess, T.R. 2002. Machining hardened steel with ceramic coated and uncoated CBN cutting tools, *Trans. NAMRI/SME* 30:337-343.
 15. Goh, G.K.L., Lim, L.C., Rahman, M., Lim, S.C. 1996. Transitions in wear mechanisms of alumina cutting tools, *Wear* 201:199–208.
 16. Brandt, G. 1986. Flank and crater wear mechanisms of alumina-based cutting tools when machining steel, *Wear* 112:39–56.
 17. Tönshoff, H.K., Bussmann, W., Stanske, C. 1986. Requirements on tool and machines when machining hard materials, *Proc. Of 26 th IMTDR Conf.*, p. 349.
 18. Stjernberg, K. and Thelin, A. 1985. In *Proc. Int. Conf. High Productivity Machining, Materials and Processes*, ASM International, 95–104.

New Correlation between Karman Constant (κ) and y -Intercept (B) for the Use of Clauser-Chart Technique to Estimate Wall Shear Stress

Sutardi¹

¹ Mechanical Engineering Department, Faculty of Industrial Technology
ITS, Surabaya, Indonesia, 60111
Tel./Fac.: +62-31-5922941, Email: sutardi@me.its.ac.id

Abstract

One of indirect wall shear stress (τ_w) measurement techniques is the use of Clauser-chart. The Clauser-chart (log-law) technique can be used reliably to estimate τ_w at higher Reynolds number flows. The accuracy of the Clauser-chart technique to estimate wall shear stress depends strongly on the values of Karman constant (κ) and the y -intercept (B). Various values of Karman constant (κ) and y -intercept (B) for turbulent boundary layer and pipe flows have been proposed by several investigators. The purpose of present study is to propose a correlation between κ and B based on various sources. Also, the proposed correlation is then checked using collection of data obtained from experiments in a low-speed wind tunnel. The study was divided into two parts: literature review and experimental study. The literature review part is intended to collect the values of κ and B from previous studies. The collection of the values of κ and B was obtained from pipe/duct and boundary layer flows. The experimental study was performed in an open circuit low speed wind tunnel, which has a $0.91\text{m} \times 0.91\text{m}$ test section and is over 20m long. The maximum freestream velocity that can be achieved is approximately 15 m/s. The freestream turbulence intensity is no larger than 0.5% at all freestream velocities. Experiments were performed at two freestream velocities of 2.0 m/s and 5.5 m/s corresponding to Reynolds numbers, based on the momentum thickness at $x = 2500\text{mm}$ from the test plate leading edge, of $R_\theta = 1000$ and 3000, respectively. Based on the data in the literatures and present experimental study, a simple linear regression between κ and B is obtained and expressed as $B = 24.9\kappa - 4.93$. Present experimental results show the validity of using the aforementioned equation in the log (overlap) region, especially at higher R_θ .

Keywords:

Wall shear stress, Clauser-chart, overlap region, Karman constant.

1. Introduction

Skin friction (wall shear stress), which is dissipative in nature, plays an important role in the system performance for both external and internal flows. For example, the skin friction drag determines the overall efficiency of aircraft, high-speed vehicles, marine vessels, piping, and ducting systems. In transportation applications, overall fuel efficiency can be improved by reducing the total drag force. The required thrust is directly proportional to drag force, and for a typical civil transport aircraft, skin friction drag can contribute up to 50 percent of the total drag at cruising speed [1]. In internal flows, such as in pipelines and ducting systems, almost 100 percent of the drag is due to skin friction. A large number of investigations of skin friction drag reduction have been performed over the last five decades, because of the economic benefits. For example, for a typical long range transport aircraft, a reduction in skin friction drag of less than 5 percent can produce a considerable savings on direct operating costs. The drag reduction is directly related to the reduction in fuel consumption, and the fuel consumption contributes up to approximately 22 percent of the total direct operating cost [2].

The wall shear stress (τ_w) on a smooth-wall turbulent boundary layer can be measured using either direct or indirect techniques. An example of a direct technique is the use of floating elements, which provides a direct reading of the 'local' wall shear stress. One disadvantage of this technique, however, is that a slight misalignment of the sensor surface will result in a large error. In addition, the presence of the gap around the sensing element creates difficulties with the measurements [3]. There are several indirect τ_w measurement techniques, including the Clauser-chart technique, power-law technique, slope of the mean velocity at the wall ($(dU/dy)_{y=0}$), and Preston tube measurements.

The Clauser-chart (log-law) technique can be used reliably to estimate wall shear stress (τ_w) at higher R_θ . The use of this technique at low Reynolds numbers, however, is more tenuous. Both DNS [4] and experimental [5] data indicate that the log region is very narrow at low R_θ . The power-law technique can provide a good estimate of τ_w in low Reynolds numbers ($R_\theta \leq 2000$, see for example, [6]). At low Reynolds numbers, the power-law represents the mean velocity in the overlap region accurately. At higher R_θ ($R_\theta \geq 3000$), on the other hand, the power-law fit in the overlap

region is poor, and it is difficult to obtain a good fit to the experimental data using the power-law approximation.

The wall shear stress (τ_w) can also be estimated from $(dU/dy)_{y=0}$. There are several difficulties, however, associated with this method. This technique needs accurate data very close to the wall i.e. $y^+ \leq 3.0$. Djenidi and Antonia [7], using LDA, showed that friction velocity (u_τ) can be determined using this technique to an accuracy of order ± 3 percent at low Reynolds numbers ($R_\theta \leq 1320$). Due to the effect of heat conduction from the wire to the wall, measurements in this region are very difficult with hot-wire anemometry, especially at high R_θ , where the linear sub-layer is very thin.

At the higher R_θ , τ_w can also be estimated using a Preston tube. This method, however, requires a high-resolution differential pressure transducer. Next, this method is usually not feasible at the lower R_θ as the reading will be out of range.

The Clauser-chart technique is still a powerful tool to determine wall shear stress (τ_w). Clauser [8] suggested following relationship for U/U_0

$$\frac{U}{U_0} = \frac{u_\tau}{U_0} g \left(\frac{u_\tau}{U_0} \frac{U_0 y}{\nu} \right) \quad (1)$$

where U_0 is the freestream velocity and u_τ is the friction velocity; y and ν are the distance from the wall and fluid kinematic viscosity, respectively.

The relation between the wall shear stress (τ_w), and the friction velocity (u_τ) can be expressed as:

$$\tau_w = \rho (u_\tau)^2 \quad (2)$$

Next, the skin friction coefficient (C_f) is defined as:

$$C_f = \frac{\tau_w}{0.5 \rho U_0^2} \quad (3)$$

where ρ is the fluid density.

The accuracy of the Clauser-chart technique to estimate wall shear stress depends strongly on the values of Karman constant (κ) and the y -intercept (B). As the author knowledge, so far there is no single value of κ and B . Various values of Karman constant (κ) and y -intercept (B) for turbulent boundary layer and pipe flows have been proposed by several investigators (Table 1). Up to now, an explicit correlation between κ and B is not available in various literatures. From Table 1, it is seen that there is dependency of κ to the value of B . As the value of κ increases, the value of B also increases. Therefore, this dependency is necessarily to be sought.

The purpose of present study is to propose a correlation between κ and B based on various sources. Also, the

proposed correlation is then checked using collection of data obtained from experiments in a low-speed wind tunnel. The proposed correlation will become a useful tool as a guidance to determine B if the value of κ is already determined, and vice versa.

2. Research Procedures

The research was divided into two parts: literature review and experimental study. Realizing that there has been large variety in the values of κ and B , the author was interested to search from many sources about the use of the Clauser-chart. This was intended to collect the values of κ and B that have been used by previous authors. The collection of the values of κ and B has been summarized, and is presented in Table 1. Table 1 shows that the types of flow, by which the values of κ and B obtained, are divided into two categories: pipe/duct and boundary layer flows.

Based on the data shown in Table 1, the author was interested to investigate more detail about the values of κ and B in the boundary layer flow using a low speed wind tunnel. Detailed experimental apparatus of current study can be found in Sutardi [26]. Following is brief description of the experimental study that had been conducted.

The experiments were performed in an open circuit low speed wind tunnel, which has a $0.91\text{m} \times 0.91\text{m}$ test section and is over 20m long, allowing investigation of relatively thick boundary layers. This alleviates to some extent the spatial resolution problems of hot-wire anemometry. The maximum freestream velocity that can be achieved is approximately 15 m/s. The freestream turbulence intensity is no larger than 0.5% at all freestream velocities. Experiments were performed at two freestream velocities of 2.0 m/s and 5.5 m/s corresponding to Reynolds numbers, based on the momentum thickness just upstream of the groove, of $R_\theta = 1000$ and 3000, respectively. The boundary layer was tripped at the leading edge of the plate using a roughness strip consisting of a piece of 100mm wide sandpaper (series 0811) and a 1.5 mm diameter cylindrical rod.

The turbulence measurements were obtained using single hot-wire anemometry. A DANTEC 55P05 single-wire boundary layer type probe connected to a DANTEC 55M01 standard bridge was used for the boundary layer measurements. The sensor is a $5\mu\text{m}$ diameter (~ 0.03 wall unit) Platinum-plated tungsten (Pt-plated tungsten) wire with an effective (active) length (l_e) of 1.25mm. In terms of wall units (ν/u_τ), the wire length is approximately 8 and 18 wall units, respectively, at $R_\theta = 1000$ and 3000. The ends of the wire are copper and gold plated to a diameter of approximately $30\mu\text{m}$. The wire was operated at an overheat ratio of 1.5. The frequency response of the wire was about 30 kHz as determined from a standard square-wave test.

All measurement uncertainties are analyzed based on the method outlined by [29], [30] and [31].

3. Results and Discussion

The friction velocity (u_τ) was determined by plotting U/U_0 versus $U_0 y/\nu$ on semi-logarithmic axes. A typical plot of the Clauser-chart for the present data at $R_\theta =$

2470 is shown in Fig. 1. The u_τ is determined by obtaining a best fit to the data, and the most appropriate value for u_τ in this case is 0.2195 m/s ($C_f = 0.003106$). This value of u_τ is obtained by trial and error. The straight lines in Fig. 1 have the form of

$$\frac{U}{U_0} = \frac{u_\tau}{U_0} \frac{1}{\kappa} \ln \frac{U_0 y}{\nu} \frac{u_\tau}{U_0} + B \tag{4}$$

Equation (4) can be expressed as

$$\frac{U}{u_\tau} = \frac{1}{\kappa} \ln \frac{u_\tau y}{\nu} + B, \tag{5}$$

where $U/u_\tau \equiv U^+$ and $u_\tau y/\nu \equiv y^+$.

Table 1. Various proposed values of κ and B .

κ	B	Proposed/used by	Flow type
0.417	5.89	Nikuradze [9]	1)
0.410	4.90	Clauser [8]	N/A
0.400	5.10	Coles [10]	N/A
0.400	5.10	Monin and Yaglom [11]	N/A
0.400-0.410	4.90-5.50	Cebeci and Smith [12]	2)
0.385 ± 10%	-	Reynolds [13]	N/A
0.410	5.00	Dean [14]	2)
0.410	5.17	Dean [14]	3)
0.400	5.50	Schlichting [15]	1)
0.394	5.56	Schlichting [15]	2)
0.410	5.00	Azad and Burhanuddin [16]	2)
0.410	5.00	Liakopoulos [17]	2)
0.400	5.00-5.50	Sill [18]	N/A
0.410	5.00	White [19]	N/A
0.400	5.00	Munson et. al [20]	1)
0.410	5.00	Bisset and Antonia [21]	N/A
0.420	5.45	Choi and Fujisawa [22]	2)
0.400	5.00	Fox and McDonald [23]	1)
0.436	6.13	Zagarolla and Smits [24]	1)
0.410	4.90	Potter and Wiggert [25]	2)
0.440	6.13	Sutardi [26]	2)
0.380	4.10	Österlund et. al [27]	2)
0.400	5.50	Fischer, Jovanovic and Durst [28]	N/A

Notes:

- N/A : Not available
- 1) : Pipe flow
- 2) : Boundary layer flow
- 3) : Duct/pipe flow

The velocity profile in Fig. 1 is re-plotted in Fig. 2 according to Eq. (5) using u_τ obtained from the best fit to the

data. The figure shows that the log-law describes the velocity well in the region $20 \leq y^+ \leq 300$ in this case. At low R_θ , the use of the log-law to estimate wall shear stress is tenuous, since the log-region at low R_θ is very narrow, or even non-existent [4][5]. In the study of [26], however, the u_τ obtained from the log-law technique compares favorably with that obtained from the power-law technique at low R_θ .

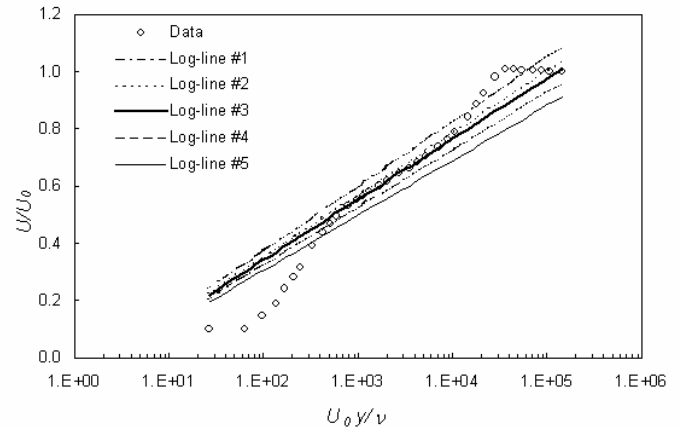


Fig. 1. Typical of the Clauser-chart for determination of u_τ . Log-line #1 is the upper most line, while log-line #5 is the lower most line. Log-line #1, $u_\tau = 0.2350$; Log-line #2, $u_\tau = 0.2250$; Log-line #3, $u_\tau = 0.2195$; Log-line #4, $u_\tau = 0.2100$; Log-line #5, $u_\tau = 0.2000$; u_τ is in m/s.

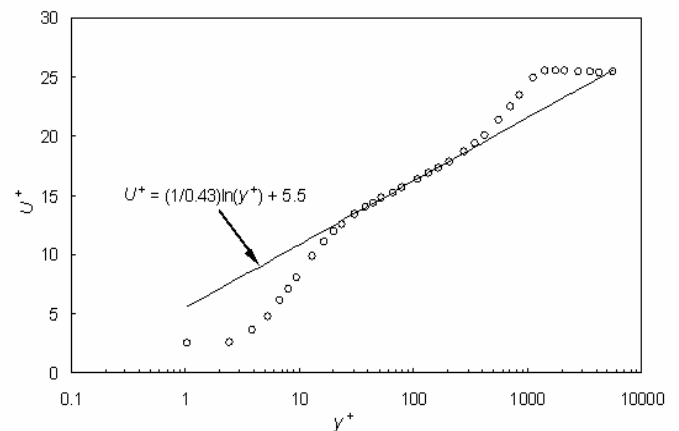


Fig. 2. Clauser-chart technique for determining u_τ presented in a $U^+ - y^+$ plane. O, smooth-wall data at $R_\theta = 2470$; —, Log-line: $U^+ = (1/0.43) \ln(y^+) + 5.5$.

The normalized mean velocity profiles obtained at $R_\theta = 1030 - 4200$ are presented in Fig. 3. The experimental profiles at the lower R_θ are in good agreement with the DNS data [4]. Very close to the wall, the experimental data deviates from the DNS data due to the wall conduction effect, and uncertainty of the spatial location of the probe [6]. In the wake region, a small deviation of the experimental data from the DNS data is discernible due to the difference in the Reynolds number. The U^+ profiles at the higher R_θ agree well with the DNS profile ($R_\theta = 1410$) up to $y^+ \approx 250$, and then deviate from the DNS data due to the difference in Reynolds

number. The U^+ profiles obtained from LDV [32] at $R_\theta = 2900$ and from HWA [33] at $R_\theta = 1840$ are also in good agreement with the present data except very close to the wall and at the outermost region (Fig. 3). In average, the uncertainty of the Clauser-chart to estimate skin friction coefficient (C_f) in the present study is approximately 5.1%.

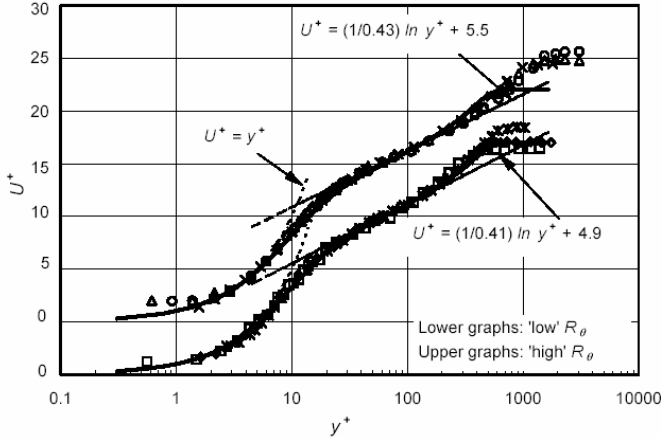


Fig. 3. Mean velocity profiles over a smooth-wall. Symbols: \square , $R_\theta = 1030$; \diamond , 1250; Δ , 3200; \circ , 4200; \times , [32], $R_\theta = 2900$; $*$, [33], $R_\theta = 1840$; —, DNS [4], $R_\theta = 1410$.

Data in Table 1 together with present experimental data for κ versus B are plotted in Fig. 4. A simple linear regression is shown in the figure together with $\pm 8\%$ error lines. The correlation between κ and B can then expressed as:

$$B = 24.9\kappa - 4.93. \tag{6}$$

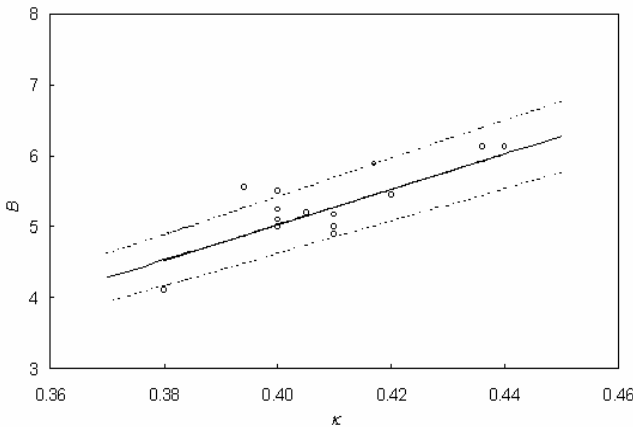


Fig. 4. Plot of the y-intercept (B) versus Karman constant (κ) from various sources in Table 1. —, linear regression line: $B = 24.9\kappa - 4.93$; ---, $\pm 8\%$ error lines.

As the author knowledge, this expression has not been proposed in any available literature. Therefore, the advantage of the use of Eq. (6) to obtain B is obvious when the value of κ is already chosen.

As shown by [34], the equation in the log (overlap) region can also be expressed as:

$$\frac{U}{u_\tau} = A \log_{10} \frac{u_\tau y}{\nu} + B, \tag{7}$$

where $A = \ln(10)/\kappa$. Using the value of $A = 5.6$ as proposed by [8], then $\kappa = 0.41$. Finally, expressing eq. (6) in term of A , one can obtain:

$$B = \frac{57.33}{A} - 4.93. \tag{8}$$

In Fig. 5, it is shown several normalized mean velocity profiles developing in a turbulent boundary layer as obtained by [26]. From the figure it can be clearly identified the validity of using eq. (5) in the log (overlap) region. All friction velocity (u_τ) value were obtained from the Clauser-chart technique as described in the beginning.

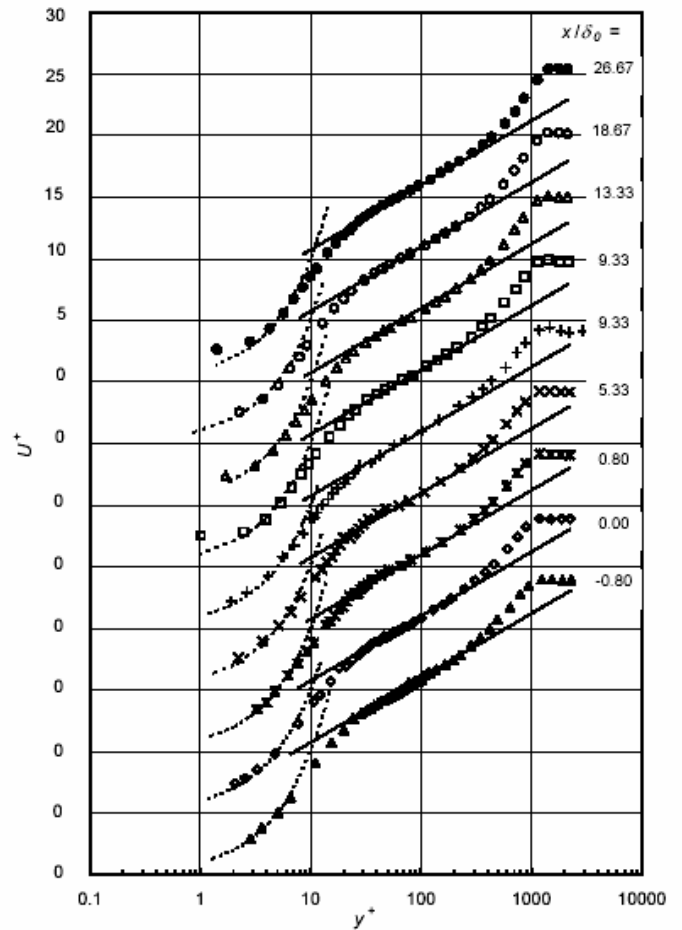


Fig. 5. Mean velocity profiles in the overlap region of a turbulent boundary layer. \blacktriangle , $R_\theta = 2840$; \diamond , 2870; $*$, 2980; \times , 3050; $+$, 3090; \square , 3100; Δ , 3200; \circ , 3400; \bullet , 4050; —, Log-law: $U^+ = (1/0.43) \ln(y^+) + 5.50$; ---, $U^+ = y^+$.

4. Conclusion

The wall shear stress (τ_w) can be accurately estimated using the Clauser-chart technique. From the available data in the literatures, there has been large variety in the values of

Karman constant (κ) and y -intercept (B). The present study proposes a new correlation of B and κ based on the available data and on the recent experimental data, and the correlation is expressed as shown in Eq. (6). Also, all claims for B and κ values, as shown in Table 1, are valid within an acceptable error range.

References

- [1] Coustols, E. and Savill, A. M. 1991. Turbulent skin friction drag reduction by active and passive means. Special Course on Skin Friction Drag Reduction. *AGARD Report 786*: 8.1-8.80.
- [2] Marec, J. P. 2000. Drag reduction: A major task for research. In Proceedings of the *CEAS/DragNet European Drag Reduction Conference 2000* 17-27. Postdam, Germany, June 19-21.
- [3] Hanratty, T. J. and Campbell, J. A. 1983. Measurement of wall shear stress. In *Fluid Mechanics Measurements* (Ed.: R. J. Goldstein), New York: Hemisphere. 559-615.
- [4] Spalart, P. R. 1988. Direct Simulation of a Turbulent Boundary Layer up to $R_\theta = 1410$. *J. Fluid Mech* 187: 61-98.
- [5] Ching, C. Y., Djenidi, L., and Antonia, R. A. 1995. Low-Reynolds-number effects in a turbulent boundary layer. *Experiments in Fluids* 19: 61-68.
- [6] Djenidi, L., Dubief, Y., and Antonia, R. A. 1997. Advantages of using a power law in a low R_θ turbulent boundary layer. *Experiments in Fluids* 22: 348-350.
- [7] Djenidi, L. and Antonia, R. A. 1993. LDA measurements in low Reynolds number turbulent boundary layer. *Experiments in Fluids* 14: 280-288.
- [8] Clauser, F. H. 1956. The turbulent boundary layer. *Advances in Applied Mechanics* 4: 1-51.
- [9] Nikuradse, J. 1933. Laws of flows in rough pipes. *NACA TM*, No. 1292.
- [10] Coles, D. 1956. The law of the wake in the turbulent boundary layer. *J. Fluid Mech* 1: 191-226.
- [11] Monin, A. S., and Yaglom, A. M. 1971. *Statistical Fluid Mechanics: Mechanics of Turbulence*. Vol. 1, Cambridge, Massachusetts: The MIT Press.
- [12] Cebeci, T., and Smith, A. M. O. 1974. *Analysis of Turbulent Boundary Layers*. 1st edition, New York.: Academic Press.
- [13] Reynolds, A. J. 1974. *Turbulent Flows in Engineering*. 1st edition, London: John Wiley & Sons.
- [14] Dean, R. B. 1976. A single formula for the complete velocity profile in a turbulent boundary layer. *J. Fluids Engineering* 98: 723-727.
- [15] Schlichting, H. 1979. *Boundary Layer Theory*. 7th edition, New York: McGraw-Hill.
- [16] Azad, R. S. and Burhanuddin, S. 1983. Measurements of some features of turbulence in wall-proximity. *Experiments in Fluids* 1: 149-160.
- [17] Liakopoulos, A. 1983. Explicit representations of the complete velocity profile in a turbulent boundary layer. *AIAA Journal* 22: 844-846.
- [18] Sill, B. L. 1986. Velocity profiles in the turbulent boundary layers. In *Advancements in Aerodynamics, Fluid Mechanics, and Hydraulics* (Eds: R. E. A. Arndt, H. G. Stefan, C. Farell, and S. M. Peterson). New York: ASCE. 9-18.
- [19] White, F. M. 1994. *Fluid Mechanics*. 3rd edition, New York: McGraw-Hill, Inc..
- [20] Munson, B. R., Young, D. F., and Okiishi, T. H. 1998. *Fundamentals of Fluid Mechanics*. 3rd edition, New York: John Wiley & Sons.
- [21] Bisset, D. K. and Antonia, R. A. 1991. Technical Note - Mean velocity and Reynolds shear stress in a turbulent boundary layer at low Reynolds numbers. *Aeronautical Journal* 95: 244-247.
- [22] Choi, K. -S. and Fujisawa, N. 1993. Possibility of drag reduction using d -type roughness. *Applied Scientific Research* 50: 315-324.
- [23] Fox, R. W., and McDonald, A. T. 1998. *Introduction to Fluid Mechanics*. 5th edition, New York: John Wiley and Sons, Inc..
- [24] Zagarola, M. V. and Smits, A. J. 1997. Scaling of the mean velocity profile for turbulent pipe flow. *Phys. Rev. Letters* 78: 239-242.
- [25] Potter, M. C., and Wiggert, D. C. 1997. *Mechanics of Fluids*. 2nd edition, Upper Saddle River, NJ.: Prentice-Hall.
- [26] Sutardi. 2003. Effect of Different Shaped Transverse Grooves on a Zero Pressure Gradient Turbulent Boundary Layer. PhD Dissertation, Fac. of Engineering and Applied Science, Memorial Univ., St. Johns, Canada.
- [27] Österlund, J. M., Johansson, A. V., and Nagib, H. M. 2000. Comment on "A note on the intermediate region in turbulent boundary layers". *Phys. Fluids* 12: 2360-2363.
- [28] Fisher, M., Jovanović, J., and Durst, F. 2001. Reynolds number effects in the near-wall region of turbulent channel flows. *Phys. Fluids* 13: 1755-1767.
- [29] Yavuzkurt, S. 1984. A guide to uncertainty analysis of hot-wire data. *J. Fluids Engineering* 106: 181-186.
- [30] Moffat, R. J. 1982. Contributions to the theory of single-sample uncertainty analysis. *J. Fluids Engineering* 104: 250-260.
- [31] Coleman, H. W., and Steele Jr., G. W. 1989. *Experimentation and Uncertainty Analysis for Engineers*. New York: John Wiley & Sons.
- [32] DeGraaff, D. B. and Eaton, J. K. 2000. Reynolds-number scaling of the flat-plate turbulent boundary layer. *J. Fluid Mech* 422: 113-135.
- [33] Purtell, L. P., Klebanoff, P. S., and Buckley, F. T. 1981. Turbulent boundary layer at low Reynolds number. *Phys. Fluids* 24: 802-811.
- [34] Schetz, J. A. 1993. *Boundary Layer Analysis*, 1st edition, New Jersey: Prentice-Hall Inc..

Cutting Force Predictions Models in End Milling Titanium Alloy Ti-6Al-4V

A.S. Mohruni^{1*2}, S. Sharif², M.Y. Noordin²

¹ Department of Mechanical Engineering, Faculty of Engineering
Sriwijaya University, 30662 Indralaya, South Sumatra Indonesia
Tel: +60-7-5534850, Fax: +60-7-5566159, E-mail: mohrunias@yahoo.com.

² Department of Manufacturing and Industrial Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5534850, Fax: +60-7-5566159, E-mail: safian@fkm.utm.my

Abstract

This paper presents a study of the development of predicted mathematical models for average tangential cutting force in end milling titanium alloy Ti-6Al-4V using uncoated solid carbide tools under flood conditions. In developing the cutting force models, the primary machining parameters such as cutting speed, feed and radial rake angle, were used as independent variables for factorial design of experiment coupled with response surface methodology (RSM). Results from the 3D-response surface contour showed that an almost constant level of cutting force was obtained during machining this advanced material. An optimum cutting conditions was also identified for a particular range of cutting force values. The models were tested by analysis of variances and were found to be adequate.

Keywords: Cutting force, End milling, Titanium Alloys, Factorial design, RSM.

1 Introduction

Titanium and its alloys are used extensively in the aerospace industry for turbine and compressor blades in the cooler parts of the engine. They are known to have excellent strength to weight ratios and corrosion resistance coupled with good elevated temperatures properties and an oxidation limit of $\sim 600^\circ\text{C}$. The α - β alloy, Ti-6Al-4V is the most common and accounts for over half of the world's sales of titanium alloys.

Numerous studies have shown titanium and its alloys are difficult to machine, regardless of the various types of cutting tools used. This has been attributed to their low thermal conductivity, which concentrates heat in the cutting zone (typically less than 25% that of steel), retention of strength at elevated temperatures and high chemical affinity for all cutting tool materials.

Although the cutting forces generated are not excessively high (almost similar to those with steel), they are confined to a small area due to the short chip contact length which leads to high stresses. The combination of high stress and temperature resulted in plastic deformation of the tool edge. Depth of cut notching and chipping at the flank can also be a problem with intermittent cutting operations. [1]

End milling is one of the most widely used machining operation and the aerospace industry places heavy demand on this process due to both the shape and complexity of the parts and the dimensional accuracy required. Recent

approaches to the problem of designing a suitable data selection system for Computer Integrated Manufacturing (CIM) application are to use machinability database systems in the form of mathematical model which have considerable advantages over simple data retrieval systems [2]. For this purpose, an approach to develop a mathematical model for the average tangential cutting force in end milling Ti-6Al-4V by factorial design of experiment coupled with RSM was conducted.

2 Cutting Forces in End Milling

The basic geometry of the end milling process for down milling is presented in Figure 1. The cutting force components acting on one tooth of the end mill cutter are shown in Figure 2. There are two cutting force components system. The first is the table system (F_x, F_y, F_z, F_R), its indices illustrated the direction of the cutting force in x-y-z coordinate respectively and the resultant force. The second is known as the cutter system of cutting forces which consists of four components F_t, F_r, F_a and F_R' , they are tangential, radial, axial and projection of the resultant force respectively [3][4].

* Corresponding Author. E-mail: mohrunias@yahoo.com,
Tel: +60-7-5534770, Fax: +60-7-5566159

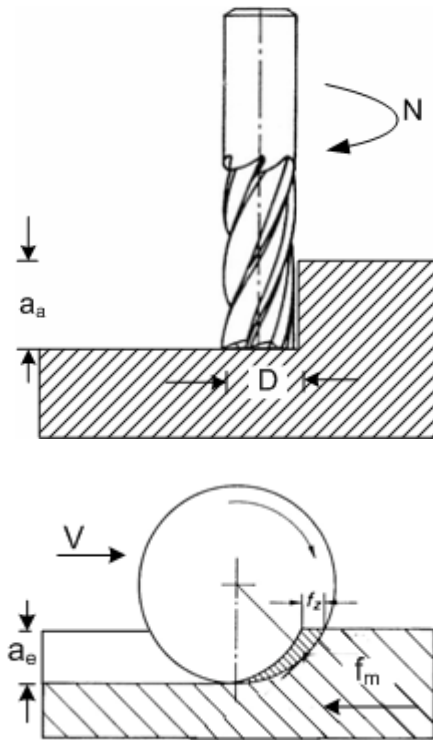


Figure 1 Basic Geometry of an End Milling Process (Down Milling)

2.1 Average Cutting Force in multi-tooth End Milling

The average cutting force is very useful for engineers in designing machine tools and in setting up the cutting system, although it is not the maximum cutting force occurred in an end milling process. If several teeth are cutting simultaneously, then the total average cutting forces acting on the teeth of the cutter per cut in table system are

$$F_{XT} = \sum_{i=1}^{z_c} \delta(i) \cdot F_{xi}(\Psi_i) \tag{1}$$

$$F_{YT} = \sum_{i=1}^{z_c} \delta(i) \cdot F_{yi}(\Psi_i) \tag{2}$$

$$F_{ZT} = \sum_{i=1}^{z_c} \delta(i) \cdot F_{zi}(\Psi_i) \tag{3}$$

where

$$\delta(i) = 1 \quad \text{if } \Psi_1 \leq \Psi \leq \Psi_2$$

$$= 0 \quad \text{otherwise}$$

F_{XT} , F_{YT} and F_{ZT} are the total average cutting forces acting on the teeth of the cutter per cut in the X, Y, and Z direction respectively, and F_x , F_y and F_z are the instantaneous cutting force on an individual tooth per cut in X, Y and Z direction respectively while Ψ_i is the instantaneous angle of the cutter.

For a multi tooth milling cutter of uniform pitch the average cutting force components in table system per tooth are

$$F_{xa} = \frac{F_{XT}}{z_c} \tag{4}$$

$$F_{ya} = \frac{F_{YT}}{z_c} \tag{5}$$

$$F_{za} = \frac{F_{ZT}}{z_c} \tag{6}$$

where z_c is the number of teeth cutting simultaneously, z_c is not being rounded off to the nearest whole number and it can be determined as

$$z_c = \frac{z \times \Psi_s}{360} \tag{7}$$

in which z is the number of teeth in the cutter and Ψ_s is the swept angle ($\Psi_2 - \Psi_1$), which can be determined in term of cutting parameters [3][4].

In cutter system for a multi-tooth milling process, the average tangential force F_{ta} per tooth and average radial force per tooth F_{ra} are

$$F_{ta} = F_t \times z_c \tag{8}$$

$$F_{ra} = F_r \times z_c \tag{9}$$

where F_t and F_r are the instantaneous tangential and radial forces acting per tooth of the cutter per cut.

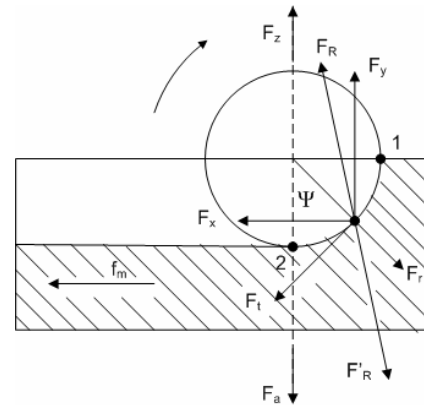


Figure 2 Cutting Force Components Acting on One Tooth of an End Mill Cutter (Down Milling).

2.2 Table and Cutter System Relationship of Cutting Force

Figure 2 shows that the average resultant cutting force acting on the workpiece in table system F_{Ra} can be determined as

$$F_{Ra} = \sqrt{F_{xa}^2 + F_{ya}^2 + F_{za}^2} \tag{10}$$

The average resultant cutting force acting on the cutter in the cutter system F_{Ra}' can be formulated as

$$F_{Ra}' = \sqrt{F_{ta}^2 + F_{ra}^2 + F_{aa}^2} \tag{11}$$

For static equilibrium it is assumed that $F_R = F_R'$ (or $F_{Ra} = F_{Ra}'$) and when the cutter is mounted correctly, the cutter axis and the spindle axis coincide each other, then it is commonly assumed that $F_z = F_a$ (or $F_{za} = F_{aa}$).

Assuming that a plane system exist, it is possible to relate the forces on the milling table to those on the cutter for down milling process as

$$F_{ta} = F_y \sin(\Psi_i) - F_x \cos(\Psi_i)$$

$$F_{ra} = F_y \cos(\Psi_i) + F_x \sin(\Psi_i) \tag{12}$$

3 Development the Mathematical Model for Cutting Forces

In machinability study investigations, statistical design of experiment is used quite extensively. Statistical design of experiment refers to the process of planning the experiment so that the appropriate data can be analyzed by statistical methods, resulting in valid and objective conclusions. Design and method such as factorial design of experiment and RSM are nowadays widely used to replace one-factor-at-a-time experimental approach which is time and cost consuming.[5]

For this purpose, the mathematical model relating to the machining response and their factor were developed to facilitate the optimization of the machining process. They have been developed stepwise using 3F1-factorial design and RSM using experimental results.

3.1 Postulation of the Mathematical Models

It is assumed that the proposed model for the cutting force is merely a function of cutting speed V , feed f_z and radial rake angle γ . Other factors such as machine tools stability, entry and exit condition etc are kept constant. Thus the proposed models for cutting force in end milling Ti-6Al-4V can be expressed as

$$F_{ta} = CV^k f_z^l \gamma^m \varepsilon' \tag{13}$$

where F_{ta} is the calculated average tangential cutting force (N), f_z is the feed per tooth (mm.tooth⁻¹), γ is the radial rake angle (°), ε' is the experimental error and C, k, l, m are parameters to be estimated using experimental data.

By performing a natural logarithmic transformation equation 13 can be converted into first order polynomial as

$$\ln F_{ta} = \ln C + k \ln V + l \ln f_z + m \ln \gamma + \ln \varepsilon' \tag{14}$$

which can also be formed as

$$y = b_0 x_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \varepsilon \tag{15}$$

and finally can be written as

$$\hat{y}_1 = y - \varepsilon = b_0 x_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 \tag{16}$$

where y is the calculated average tangential force on a natural logarithmic scale, \hat{y}_1 is the natural logarithmic value of predictive (estimated) tangential cutting force, $x_0 = 1$ (a dummy variable), x_1, x_2 and x_3 are the coded variables of V, f_z , and γ respectively, $\varepsilon = \ln \varepsilon'$ and b_0, b_1, b_2 and b_3 are the model parameters to be estimated using the experimental data. [6]

In extended observation region, the second-order model is also useful when the second order effect of V, f_z, γ and the two way interactions amongst V, f_z , and γ are significant. The second order can be extended from the first-order model in equation 16 as

$$\begin{aligned} \hat{y}_2 = y - \varepsilon &= b_0 x_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 \\ &+ b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 \\ &+ b_{11} x_1^2 + b_{22} x_2^2 + b_{33} x_3^2 \end{aligned} \tag{17}$$

where the b values are the parameters, which are to be estimated by method of least squares and \hat{y}_2 is the estimated

response on logarithmic scale.

Validity of the models used for optimizing the process parameters has to be tested using ANOVA.

3.2 Experimental Works

Before commencing the experimental trials, thorough planning was essential in order to obtain the relevant data in developing the mathematical models. By taking into consideration the factors for experimentation and analysis such as cutting speed, feed and radial rake angle, the design of experiments (DOE) were used stepwise from 2³-factorial design to central composite design (CCD), which is easily gained by augmentation 2³-design with replicated star points.

3.2.1 Experimental Design

In this study, the 2³-factorial design shown in Figure 3, was used as screening trials of the experiments. This is one of which all levels of a given factor are combined with all levels of every other factor in the experiment. This design is necessary when interactions between variables are to be investigated. Furthermore, factorial design allow the effects of a factor to be estimated at several levels of other factors, giving conclusions that are valid over a range of experimental conditions [7][8].

To observe the effect of non linearity in the region and to construct an estimate of error with $n_c - 1$, it is useful to use additional center points in screening with 2 level factorial designs when the factorial points in the designs are not replicated [6][8].

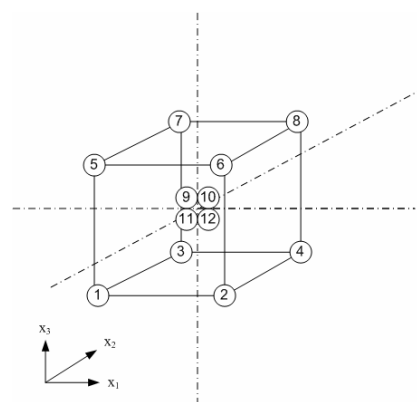


Figure 3 3F1-Factorial Design Augmented with 4 Center Points and First-Order CCD for $k = 3$.

An extended design of 2³-design is a second order CCD design, which is augmented with replicated star points as shown in Figure 4. The numbers of such repeated measurements affect the distance of the “axial star points” within the factor space. According to previous study [6] the distance of axial star points to the center points α is 1.4142.

3.2.2 Coding of the Independent Variables

The variables were coded by taking into account the capacity and limiting cutting conditions of the milling machine. The following transforming equation was used.

$$x = \frac{\ln x_n - \ln x_{n0}}{\ln x_{nl} - \ln x_{n0}} \tag{18}$$

where x is the coded variable of any factor corresponding to

its natural x_n , x_{n1} is the natural value at the +1 level and x_{n0} is the natural value of the factor corresponding to the base or zero level [2][3][6][7]. The level of the independent variables and coding identification are illustrated in Table 1.

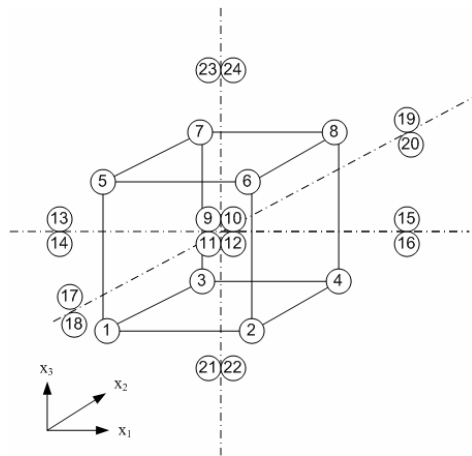


Figure 4 Second Order CCD for k = 3.

Table 1 Levels of Independent Variables for Ti-6Al-4V

Independent Variable	Level in coded form				
	-α	-1	0	+1	+α
V (m.mm ⁻¹) x ₁	124.53	130.00	144.22	160.00	167.03
f _z (mm.tooth ⁻¹) x ₂	0.025	0.03	0.046	0.07	0.083
γ (°) x ₃	6.2	7.0	9.5	13.0	14.8

3.2.3 Experimental Set-Up

For the experimentation, a CNC MHO 700S milling machine was used for side milling process, which was carried out with a constant a_a (axial depth of cut) 5 mm and a_c (radial depth of cut) 2 mm under flood conditions with a 6% concentration of water base coolant. The grade-K30 solid carbide end mill cutter with different radial rake angle according to design of experiment, were used in the experiments.

The cutting forces were measured at the first cut of the whole cutting process of each end mill cutter using multi component force measuring system consisting of the following elements:

- A 3-component dynamometer comprising of basic unit (Kistler, Type 9265B) and a screwed-on working adapter for milling (Kistler, Type 9443B).
- A multi channel charge amplifier (Kistler, Type 5019A).
- A data acquisition system consisting of a personal computer (PC) equipped with an A/D board as well as the DynoWare software (Kistler, Type 2825 D1-2, version 2.31).

The reference workpiece material of Ti6Al-4V was a rectangular block of 110 mm x 110 mm x 400 mm and the

analysis for the developed models were carried out using a Design Expert 6.0 package.

4 Experimental Results and Discussion

4.1 The 3F1-Model of the Cutting Force

Using the experimental results in Table 2, the cutting force prediction model can be formulated as

$$\hat{y} = 4.237 - 0.01052x_1 + 0.3123x_2 - 0.0546x_3 + 0.02611x_1x_3 - 0.03472x_2x_3 \tag{19}$$

This equation shows that the cutting force decreases with increasing cutting speed and radial rake angle, and in contrary it increases with increase in feed. From the interaction terms, it was observed that the combination of speed and radial rake angle contributes to the increase in cutting force. However the combination of feed and radial rake angle adversely reduces the cutting force, whilst the feed alone tends to increase in cutting force. The response surface of the cutting force distribution in relation to cutting speed and rake angle is shown graphically in Figure 5.

Table 2 Experimental Results for 3F1-Factorial and Linear CCD-Model with k = 3

Std	Type	V	f _z	γ	Calculated Fc N
		m.min ⁻¹	mm/tooth	deg.	
1	Fact	-1	-1	-1	53.66
2	Fact	1	-1	-1	49.75
3	Fact	-1	1	-1	107.16
4	Fact	1	1	-1	99.83
5	Fact	-1	-1	1	48.87
6	Fact	1	-1	1	50.45
7	Fact	-1	1	1	85.20
8	Fact	1	1	1	87.84
9	Center	0	0	0	76.70
10	Center	0	0	0	74.84
11	Center	0	0	0	73.15
12	Center	0	0	0	79.99

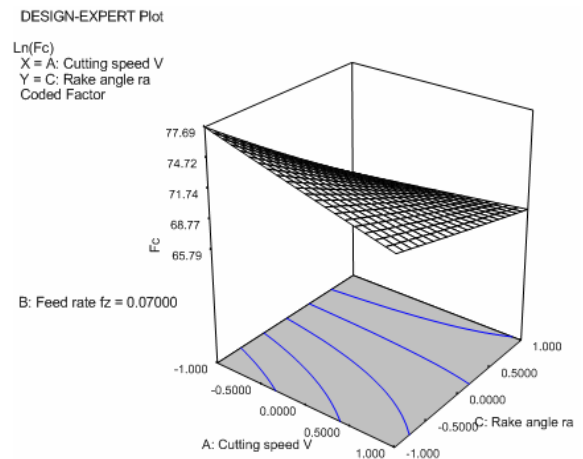


Figure 5 Response Surface for 3F1-Factorial Model

By using ANOVA, the validation of this equation is shown in Table 3. From the results, it is evident that the 3F1-model is valid for the observation region, because the lack of fit (LOF) of this model is not significant.

Table 3 ANOVA for 3F1-Factorial Model

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	significant
Model	0.8201	5	0.1640	204.6	< 0.0001	significant
A0.0008857	1	0.0008857	1.105	0.3413		
B 0.7803	1	0.7803	973.4	< 0.0001		
C 0.02385	1	0.02385	29.75	0.002816		
AC 0.005452	1	0.005452	6.801	0.04779		
BC 0.009643	1	0.009643	12.03	0.01788		
Curvature0.02732	1	0.02732	34.08	0.002086	significant	
Residual0.004008	5	0.0008016				
Lack of Fit6.192E-006	2	3.096E-006	0.002321	0.9977	not significant	
Pure Error0.004002	3	0.001334				
Cor Total	0.8515	11				

4.2 The First-Order CCD-Model

The same data in Table 2 was used for developing the first order CCD-model. The first order model for cutting force is

$$\hat{y} = 4.271 - 0.01052x_1 + 0.3123x_2 - 0.05460x_3 \quad (20)$$

Equation 20 can be presented in the following form:

$$F_c = 1711.6736V^{-0.10133} f_z^{0.73717} \gamma^{-0.1764} \quad (21)$$

The results generated from equations 19 and 20 showed that they have the same coefficient in the linear region. The difference between both equations in the linear region is merely on their intercepts, i.e 4.237 and 4.271 for 3F1 model and linear CCD model respectively. It means that same effect are obtained from both equations in the observation region, however the 3F1-factorial model provides more information about the intersection effect between the cutting speeds combined with radial rake angle and between the feed combined with radial rake angle.

More information resulted in CCD linear model is shown by the response surface in Figure 6. From this figure it can be recognized that with increasing cutting speed, the cutting force decreases very slightly. Similar finding was reported by other researchers [9][10] for the observation region of cutting speed.

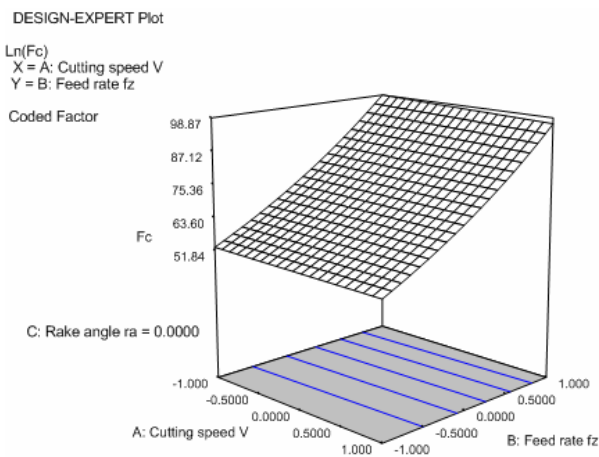


Figure 6 Response Surface for Linear CCD-Model

For validation of the linear CCD-model, ANOVA was used and the results are shown in Table 4. It shows that the LOF of the first order CCD model is not significant, thus the model can be accepted within the observation region.

Table 4 ANOVA for the First Order CCD-Model

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	significant
Model	0.8050	3	0.2683	46.24	< 0.0001	significant
A0.0008857	1	0.0008857	0.1526	0.7062		
B 0.7803	1	0.7803	134.5	< 0.0001		
C 0.02385	1	0.02385	4.110	0.07718		
Residual 0.04643	8	0.005803				
Lack of Fit0.04242	5	0.008485	6.361	0.07937	not significant	
Pure Error0.004002	3	0.001334				
Cor Total	0.8515	11				

4.3 The Second-Order CCD-Model

A second-order model was postulated to extend the variables range in obtaining the relationship between the cutting force and the machining variables. The model is based on the second order CCD for k=3 (Figure 4) and 24 set of experiments given in Table 5. The result is presented in the following form:

$$\hat{y} = 4.35 + 0.008803x_1 + 0.2587x_2 - 0.09927x_3 - 0.02905x_1^2 - 0.05502x_2^2 - 0.01716x_3^2 + 0.0004374x_1x_2 + 0.02611x_1x_3 - 0.03472x_2x_3 \quad (22)$$

Table 5 Experimental Results of the Second Order CCD-Model for k = 3

Std	Type	V	f _z	Γ	Calculated F _c N
		m.min ⁻¹	mm/tooth	Deg.	
1	Fact	-1	-1	-1	53.66
2	Fact	1	-1	-1	49.75
3	Fact	-1	1	-1	107.16
4	Fact	1	1	-1	99.83
5	Fact	-1	-1	1	48.87
6	Fact	1	-1	1	50.45
7	Fact	-1	1	1	85.20
8	Fact	1	1	1	87.84
9	Center	0	0	0	76.70
10	Center	0	0	0	74.84
11	Center	0	0	0	73.15
12	Center	0	0	0	79.99
13	Axial	-1.4142	0	0	74.84
14	Axial	-1.4142	0	0	67.53
15	Axial	1.4142	0	0	68.56
16	Axial	1.4142	0	0	86.43
17	Axial	0	-1.4142	0	51.22
18	Axial	0	-1.4142	0	53.90
19	Axial	0	1.4142	0	100.82
20	Axial	0	1.4142	0	87.41
21	Axial	0	0	-1.4142	95.33
22	Axial	0	0	-1.4142	90.45
23	Axial	0	0	1.4142	55.69
24	Axial	0	0	1.4142	68.59

It was interesting to observe that when the region was extended, the contour of cutting force in the cutting speed range changes from linear (Figure 6) to a slightly curve form (Figure 7). This was also confirmed by other researchers [9][10] for low and high cutting speeds region. They found that the cutting force was very high at low cutting speed and reduced rapidly at medium cutting speed and finally increased slightly with further increase in cutting speed. It was also observed in Figure 7 that there was a significant increase in cutting force with increase in feed.

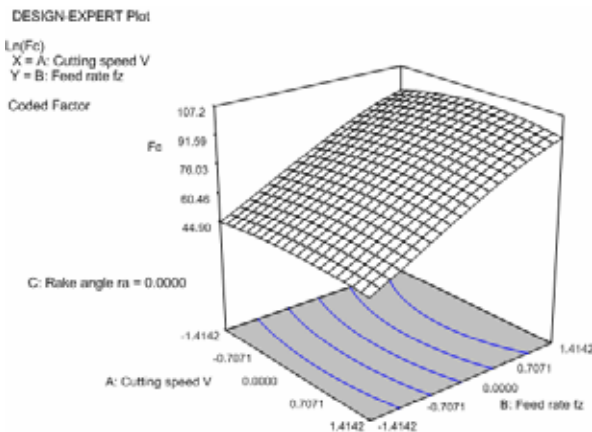


Figure 7 Response Surface for the Second Order CCD-Model

From ANOVA results, it was also found that the second order CCD model can be used as the mathematical model in the region of observation, since the LOF of this model is not significant as shown in Table 6.

Table 6 ANOVA for the Second Order CCD-Model

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	
Block	0.003271	1	0.003271			
Model	1.285	9	0.1427	12.01	< 0.0001	significant
A: 0.001240	1	1	0.001240	0.1043	0.7518	
B: 1.071	1	1	1.071	90.14	< 0.0001	
C: 0.1577	1	1	0.1577	13.27	0.002980	
A:B: 0.01012	1	1	0.01012	0.8520	0.3728	
B:C: 0.03632	1	1	0.03632	3.057	0.1040	
C:A: 0.003532	1	1	0.003532	0.2973	0.5948	
A:B:C: 1.531E-006	1	1	1.531E-006	0.0001288	0.9911	
Residual	0.1545	13	0.01188			
Lack of Fit	0.08381	4	0.02095	2.668	0.1020	not significant
Pure Error	0.07068	9	0.007853			
Cor Total	1.442	23				

5 Conclusions

- There are three appropriate prediction models namely 3F1-, linear CCD and second order CCD model to formulate the relationship amongst the machining parameters such as cutting speed (130-160 m/min), feed (0.03-0.07 mm/tooth), radial rake angle (7-13 °).
- With increasing cutting speed, the cutting force decreases slightly in the region of observation.
- Feed is the most significant factor that influences the cutting force. It increases significantly with increasing feed in the observation region.

- Increasing the radial rake angle gradually reduced the cutting force.

Acknowledgements

The authors wish to thank the research Management Center, UTM and the Ministry of Science, Technology and Innovation Malaysia for their financial support to the above project through the IRPA funding 03-02-02-0068 PR0074/03-01 – Vote no. 74545.

References

- [1]. Niemann, H.; Eu-gene Ng.; Loftus, H.; Sharman, A.; Dewes, R. and Aspinwall, D. 2002, The Effect of Cutting Environment and Tool Coating when High Speed Ball Nose End Milling titanium Alloy, In *Metal Cutting and High Speed Machining*, edited by Dudzinski, D.; Molinari, A.; Schulz, H., Kluwer Academic/Plenum Publisher.
- [2]. Alauddin, M.; El Baradie, M.A.; Hashmi, M.S.J. 1996, Modelling of Cutting Force in End Milling Inconel 718, *Journal of Material Processing Technology* 58: 100-108.
- [3]. Alauddin, M. 1993, End Milling Machinability of Steel, a Nickel-base Alloy (Inconel 718) and a Metal Matrix Composite. PhD Thesis, Dublin City University.
- [4]. Paucksch, E. 11th eds. 1996, *Zerspantechnik*, Viewegs-Fachbuecher der Technik, Braunschweig.
- [5]. Noordin, M.Y.; Venkatesh, V.C.; Sharif, S.; Elting, S.; Abdullah, A. 2004, Application of Response Surface Methodology in Describing the Performance of Coated Carbide Tools when Turning AISI 1045 Steel, *Journal of Materials Processing Technology* 145: 46–58.
- [6]. Sharif, S.; Mohruni, A.S.; Noordin, M.Y. 2006, Modeling of Tool life when End Milling on Titanium Alloy (Ti-6Al-4V) using Response Surface Methodology, In *Proceeding of the 1st International Conference & 7th AUN/SEED-Net Fieldwise Seminar on Manufacturing and Material Processing, 14-15 March: 127-132*.
- [7]. Choudhury, I.A.; El-Baradie, M.A. 1999, Machinability assessment of Inconel 718 by Factorial Design of Experiment Coupled with Response Surface Methodology, *Journal of Materials Processing Technology*, 95: 30-39.
- [8]. Meyers, R.H.; Montgomery, D.C. 2nd eds. 2002, *Response Surface Methodology: Process and Product Optimization using Designed Experiments*, John Wiley & Sons, Inc.
- [9]. Trent, E.M.; Wright, P.K. 4th eds. 2000, *Metal Cutting*, Butterworth-Heinemann
- [10]. Xu, J.H.; Ren, K.Q.; Geng, G.S. 2004, Cutting Forces in High Speed Milling of a Close Alpha Titanium Alloy, In *Key Engineering Materials Vols. 259-260: 451-455*.

Additives in EDM Dielectric

Norliana Mohd Abbas*, Darius G. Solomon and Md. Fuad Bahari

Faculty of Mechanical Engineering,
Universiti Teknologi MARA (UiTM), Shah Alam, Selangor Darul Ehsan, Malaysia
Tel: +603-5543 5159, Fax: +603-5543 5160
*Email address: chelorot@lycos.com

Abstract

Electrical Discharge Machining (EDM) has been developed in late 1940s. The process is based on removing material from a part by a series of repeated electrical discharges between tool called electrode and the work piece with the presence of dielectric. The dielectric serves as transportation of removed particles, remains electrically non-conducting until the breakdown voltage is reached, reconditions the dielectric strength, increases the energy density in plasma channel and cools the electrode. Petroleum based hydrocarbon mineral oils and deionised water are the most commonly used dielectric. Recent trends involve the use of clear and low-viscosity fluid as dielectric. Studies have found that addition of additives in dielectric leads to the increase of material removal rate (MRR), thinner recast layer and better surface finish (SF). This paper discusses the selection of dielectric fluids and the types of additives added in dielectric. Material removal mechanism and the influence of additives in MRR, recast layer and SF are explained. The usage of additives in dielectric in Malaysian industries is explored based on a survey conducted in Selangor area.

Keywords: EDM, dielectric, powder additives, liquid additives, material removal rate, tool wear, surface finish

1. INTRODUCTION

Electrical Discharge Machining (EDM) is a non-traditional concept of machining which has been widely used to produce dies and moulds. It is also used for finishing parts for aerospace and automotive industry and surgical components [1, 2, 3]. The electrode is moved toward the work piece until the gap is small enough so that the impressed voltage is great enough to ionize the dielectric [4]. Short duration discharges are generated in a liquid dielectric gap, which separates tool and work piece. The material is removed with the erosive effect of the electrical discharges from tool and work piece [5]. EDM does not make direct contact between the electrode and the work piece where it can eliminate mechanical stress, chatter and vibration problems during machining [3]. Materials of any hardness can be cut as long as the material can conduct electricity [6].

Basic EDM process consists of electrode, work piece materials, dielectric and the range of pulse rate, current and voltage. Figure 1 shows the schematic diagram of EDM process [1]. EDM electrode is a shaped tool and it is made as a mirror image of the part that is going to be produced. Materials typically used for EDM electrodes are various forms of graphite, copper, tungsten, brass, silver and steel [8]. It should possess high thermo-mechanical characteristics and retain the specified-accuracy of the final shape [9]. Most homogeneous materials such as aluminium, copper, tungsten and titanium are electrically conductive

materials and regardless of their hardness, it can be machined using EDM process [10]. The research interest on the work piece materials is towards the application of the materials in industries viz. metals, composites and conductive ceramics.

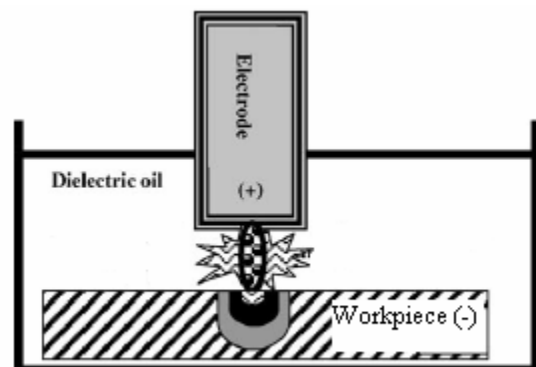


Figure 1: Schematic diagram of EDM process [1]

The dielectric is used to transport the removed particles, to increase the energy density in plasma channel, to recondition of the dielectric strength, to remain electrically non-conducting until the breakdown voltage is reached and to cool the electrode [11]. Petroleum-based hydrocarbon and deionised water are the two most commonly used dielectrics. It is pumped through the arc gap to flush away the eroded particles between the work piece and the electrode. In EDM process, the spark discharge is repeated

at rates between 50 kHz to 500 kHz. The voltage range between 50V and 380V and the current supply is from 0.1A to 500A [12].

This paper discusses the types of dielectric fluids used in EDM, types of additives added in EDM dielectric, influence of powder additives in EDM performance and EDM in Malaysian industries.

2. Dielectric Fluid

A book written by B. Kumar and Sujay Kumar [13] indicates some essential requirements of a dielectric fluid. The dielectric fluid should not be inflammable, should not evolve toxic vapours during operation and should not react with the material, tool or the container. It should be cheap, easily available and have an optimum viscosity. This is because, lower viscosity disperse the fluid to carry away the metal particles efficiently at a flow velocity while higher velocity imposes restriction on the liquid flow. Other elements to be considered in selection of dielectric are flash point, dielectric strength, specific gravity, colour and odour [14].

Petroleum based hydrocarbon oil and deionised water are the two most commonly used dielectric. Other than that, silicon fluid, mixture of silicon fluids with petroleum oil, aqueous solution of ethylene glycol and water in emulsion can also be used [14]. Synthetic EDM dielectric is also available in the market. Most synthetic fluids are manufactured from chemical constituents or polymerization of a hydrocarbon (olefins) which is more expensive but have better oxidation stability, higher flash point and less corrosive and toxic [15]. Recent trends involve the use of clear and low-viscosity fluid [12].

2.1 Oil Based Dielectric

For an oil based dielectric, kerosene as the dielectric is very common among the researchers [1,16,17] even though it is inflammable, having possibility of fire hazard and will produce solid solubilisation and diffusion of carbon on the machined surface.

H. Ramasawmy and L. Blunt [6] used BP 180 as dielectric to quantify the effect of some of the main EDM parameter on surface texture. BP 180 is recognized by Intech EDM company as the popular brand and commonly used dielectric fluids in industry. W.S. Zhao et al [18] used EDM special working fluid from Mobil in investigating the machining efficiency and surface roughness of powder mixed in EDM dielectric. M.Z. Zahiruddin et al [19] used FUCHS lubricant as dielectric when studying the effect of electrical parameters on the EDM performances of titanium alloy.

2.2 Water Based Dielectric

Water as dielectric is an alternative to hydrocarbon oil. The approach is taken to promote a better health and safety

environment while working with EDM. This is because hydrocarbon oil such as kerosene will decompose and release harmful vapour (CO and CH₄) [20]. The first paper about the usage of water as dielectric was published by Jeswani M.L. [21] in 1981. He compared the performances of kerosene and distilled water where machining in distilled water resulted in a higher MRR and a lower wear ratio than in kerosene when a high pulse energy range was used. With distilled water, the machining accuracy was poor but the surface finish was better. S. Tariq Jilani and P. C. Pandey [22] investigated the performance of water as dielectric in EDM using distilled water, tap water and a mixture of 25% tap and 75% distilled water. The best machining rates have been achieved with the tap water and machining in water has the possibility of achieving zero electrode wear when using copper tools with negative polarities.

Konig W. and Siebers F.J. [23] explained the influence of the working medium on the removal process. They indicated that working medium has a sustained influence on the removal process. The erosion process in water-based media consequently possesses higher thermal stability and much higher power input can be achieved especially under critical conditions, allowing much greater increases in the removal rate. A considerable difference between conventional oil-based dielectrics and aqueous media is specific boiling energy of aqueous media is some eight times higher and boiling phenomena occur at a lower temperature level.

Researchers may find various types of EDM dielectric product in the market. The selection of dielectric fluids mostly depends on the availability of the product and the capability of the EDM machine. In EDM, distilled water is rarely used. Although it gives high MRR and increased cooling capacity, it also gives high electrode wear. However, the greatest used of distilled water can be found in Wire EDM [24].

3. Types of Additives

Generally there are two types of additives used in EDM: powder additives and liquid additives. An electrically conductive powder is added in the oil-based dielectric with the range of 0.01 to 10 micron granularity and 2 to 80g/l powder concentration. Organic compound which is in liquid based is added in water-based dielectric to improve the machining performances.

3.1 Powder Additives

The hybrid material removal process is called powder mixed EDM (PMEDM) where it works steadily at low pulse energy [25] and it significantly affects the performance of EDM process. Electrically conductive powder reduces the insulating strength of the dielectric fluid and increase the spark gap between the tool and the work piece. EDM process becomes more stable and improves machining efficiency, MRR and SF. However, most studies were conducted to evaluate the surface finish since the

process can provide mirror surface finish which is a challenging issue in EDM. The characteristics of the powder such as the size, type and concentration influence the dielectric performance.

3.1.1 Material Removal Mechanism using Powder Additives

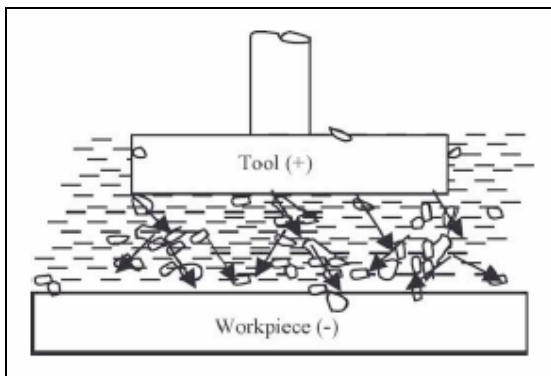


Figure 2: Principles of MRR using additives [17]

Addition of powder additives in EDM process alters the material removal mechanism which leads to an increase in gap between the tool and the work piece. The powder particles get energized and behave in a zigzag fashion (refer to Figure 2). The grains close to each other under the sparking area and gather in clusters. The interlocking between the different powder particles takes place due to variation in their shape and size. They arrange themselves in the form of chain and helps in bridging the gap between both the electrodes. The easy short circuit takes place, which causes early explosion in the gap. As a result, the series discharge starts under the electrode area. Due to increase in frequency of discharging, the faster sparking within a discharge takes place which causes faster erosion from the work piece surface.

An additional dielectric circulating system is needed to avoid settling of powder particle during machining. Magnetic force is also employed to separate the powder particles from the debris.

3.1.2 Types of Powder Additives

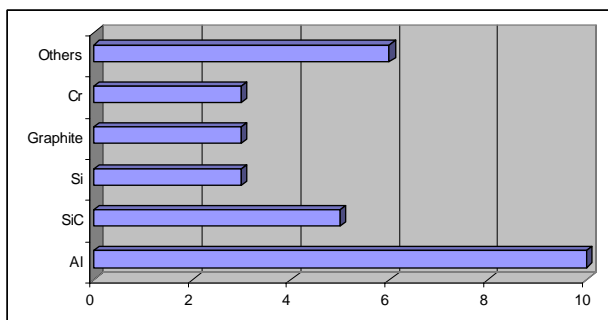


Figure 3: Distribution of type of powder used based on the collected papers

From collected literature, a bar chart is made (refer to Figure 3) to show the type of powder used and the number of works dealing with each type of powder is compared. Aluminium (Al) powder attracts many researchers since it contributes well in improving MRR as well as surface finish. Silicon carbide (SiC) has been used in five experiments while silicon (Si), graphite (Gr) and chromium has been used in three experiments respectively. Molybdenum sulphide, crushed glass and copper powder is categorised as others. Copper powder in kerosene makes almost no difference.

Yan B.H et al [26] revealed that the corrosion resistance and surface hardness were improved by adding the proper powder into dielectric. It is important to have the correct combination of powder and work piece materials and an understanding of the fundamental mechanisms affecting such combinations will promote the applications of PMEDM to feasibly produce superior surface finish and properties of components using EDM [27].

3.2 Liquid Additives

F.N. Leao and I.R. Pashby [28] found that some researchers have studied the feasibility of adding organic compound such as ethylene glycol, polyethylene glycol 200, polyethylene glycol 400, polyethylene glycol 600, dextrose and sucrose to improve the performance of deionized water. Koenig W. and Joerres L. [29] reported that a highly concentrated aqueous glycerine solution has an advantage as compared to hydrocarbon dielectrics when working with long pulse durations and high pulse duty factors and discharge currents, i.e. in the roughing range with high open-circuit voltages and positive polarity tool electrode. The surface of titanium has been modified after EDM using dielectric of urea solution in water [30]. The nitrogen element decomposed from the dielectric that contained urea, migrated to the work piece forming a TiN hard layer which resulting in good wear resistance of the machined surface after EDM.

Major research studies are concerning more in using powder as additives. Not much information is gathered about the liquid additives and the mechanism of the material removal using liquid additives.

4. Influence of Powder Additives in EDM Performance

As proved by experiments, addition of powder additives can improve SF, MRR and low recast layer.

Silicon powder was used by P.Pecas and E. Henriques [31] to assess improvement through quality surface indicators and process time management over a set of different processing area. The result shows that 2g/l of Si concentration, smooth and high reflective craters were achieved with average surface roughness (R_a) depends on the area and varies between 0.09 μm for 1 cm^2 and 0.57 μm for 64 cm^2 electrode. T. Yih-fong and C. Fu-chen [32]

investigates the effect of powder properties on surface quality of SKD-11 work piece using Al, chromium (Cr), copper (Cu), and SiC powders. The smallest particle (70–80 nm) generates best surface finish and Al powder produces the best surface finish.

Machining efficiency and surface roughness of rough PMEDM in rough machining was studied by W.S. Zhao et al [18] using Al with 40g/l and 10 μ m granularity and they discovered that machining efficiency was improved from 2.06 to 3.4mm³/min with an increasing rate of 70%. The machining efficiency can be highly increased by selecting proper discharge parameter (increasing peak current and reducing pulse width) with better surface finish in comparison with conventional EDM machining.

Influence of powder particle in micro EDM especially in the gap, thermal spread in dielectric and influenced zone was investigated on 718 inconel alloy by F. Klocke et al [33]. Al and Si powder was added with 2g/l and 10g/l concentration respectively. The result shows that Al powder leads to thinnest rim zone and highest MRR. K.L Wu et al [16] added surfactant with Al powder and found that it has more apparent effect to distribute electrical discharging energy to achieve thinner recast layer around 1–2 μ m in thickness compared to 5–8 μ m under pure dielectric.

5. EDM in Malaysian Industries

A survey was carried out in a few industries in Selangor area. The survey was done through interviews with the employees who are in charge of EDM machines. The questions were formulated from the problems observed through literature review; questions were based on current practices, products made and problems faced by industries. Most of the products produced by our local industries using EDM are moulds and dies for injection moulding, metal stamping, electrical and electronic industries. Some of the companies produce finish parts for automotive, defence and telecommunication industries.

The important elements evaluated by industries for the product produced using EDM are dimension, surface finish and the machining time. The surface finish is measured using templates. There is a standard template used to check the SQ based on the requirement of the customer. The selection of the electrical parameters will be based from the coding of the template and the machine instruction manual. Not much further inspection is done in order to evaluate the surface of products made through EDM. Some of the industries used lens or microscope to verify the smoothness of the surface. Other than that, they conduct trial tests on the mould or dies produced through EDM. If the outputs of the trial tests meet the customer requirements, there will be no further evaluation on the products.

Kerosene is widely used as dielectric in Malaysian industries even though they claimed that it is smelly and produced smoke. Besides kerosene, they used refined

petroleum such as HONELO 420, CASTROL, FUCHS and Petro Canada 125. The selection of different dielectric is to get better surface finish and safety (fire proof), to prevent skin irritation and bad smell. None of the industries implement the technique of using additives in dielectric in order to increase the SQ, TW and MRR. Some of the industries undergo polishing and chroming in order to get better SQ for the final part.

6. Discussions

This paper presents a general overview about additives in EDM machining, the influence of additives in EDM performance and the application of the technique in Malaysian industries. Many investigations have been made on the usage of powder additives but there is a lack of information on the usage of liquid additives in oil-based dielectric.

7. Acknowledgements

The authors would like to express the gratitude to:

- a) Universiti Teknologi MARA for funding the current research studentship under the Young Lecturer Scheme
- b) The management and the staff of Innopeak (M) Sdn. Bhd., Moltec Precision Sdn. Bhd., Solsius Technology Sdn. Bhd., HICOM Diecastings Sdn. Bhd., High Advantage Sdn. Bhd., Ejra Carbide Sdn. Bhd., MIC-CO Sdn. Bhd. and Polymatech (M'sia) Sdn. Bhd. for the information given.

8. References

- [1] Katsushi Furutani; Akinori Saneto; Hideki Takezawa; Naotake Mohri and Hidetaka Miyake. 2001. Accretion of titanium carbide by electrical discharge machining with powder suspended in working fluid. *Journal of the International Societies for Precision Engineering and Nanotechnology* 25:138–144.
- [2] J. Simao; H.G. Lee; D.K. Aspinwall; R.C. Dewes and E.M. Aspinwall. 2003. Workpiece surface modification using electrical discharge machining. *International Journal of Machine Tools & Manufacture* 43:121–128.
- [3] K.H. Ho and S.T. Newman. 2003. State of the art electrical discharge machining (EDM). *International Journal of Machine Tools & Manufacture* 43:1287–1300.
- [4] B. Bojorquez; R.T. Marloth and O.S. Es-Said. 2002. Formation of a crater in the workpiece on an electrical discharge machine. *Engineering Failure Analysis* 9:93–97.
- [5] J. Marafona and J.A.G. Chousal. 2005. A finite element model of EDM based on the Joule effect. *International Journal of Machine Tools & Manufacture* xx:1–8.

- [6] H. Ramasawmy and L. Blunt. 2004. Effect of EDM process parameters on 3D surface topography. *Journal of Materials Processing Technology* 148:155–164.
- [8] Li Li; Y. S. Wong; J.Y.H. Fuh and L. Lu. 2001. EDM performance of TiC_copper-based sintered electrodes. *Materials and Design* 22:669-678.
- [9] Frederic Gillot; Pascal Mongol and Benoit Furet. 2005. Dimensional accuracy studies of copper shells used for electro-discharge machining electrodes made with rapid prototyping and the electroforming process. *Journal of Materials Processing Technology* 159:33–39.
- [10] Eric Oberg, Franklin D. Jones, Halbrook L. Horton and Henty H. Ryffel. 2004. *Machinery's Handbook 27th Edition*. Industrial Press Inc., New York.
- [11] K.P. Rajurkar. Edited by Richard C. Dorf and Andrew Kusiak. 1994. *Nontraditional Manufacturing Processes, Handbook of Design, Manufacturing and Automation*. John Wiley and Sons, USA.
- [12] Serope Kalpakjian and Steven R. Schmid. 2001. *Manufacturing Engineering and Technology Fourth Edition*. Prentice Hall International, USA.
- [13] B. Kumar and Sujay Kumar. 2002. *Manufacturing Processes and Technology, Fourth Edition*. Khanna Publisher India.
- [14] Dielectric fluid., 18 December 2005
Citing internet sources URL
<http://www.unl.edu/nmrc/Diesinking/surfaceint/surface.htm>
- [15] Everything you need to know about dielectric fluids, 19 October 2005. Citing internet sources URL
<http://www.intech-edm.com/pdf/spksF100.pdf/date>
downloaded
- [16] Kun Ling Wu; Biing Hwa Yan; Fuang Yuan Huang and Shin Chang Chen. 2005. Improvement of surface finish on SKD steel using electro-discharge machining with aluminum and surfactant added dielectric. *International Journal of Machine Tools & Manufacture* 45:1195–1201.
- [17] H.K. Kansal; Sehijpal Singh and P. Kumar. 2005. Parametric optimization of powder mixed electrical discharge machining by response surface methodology. *Journal of Materials Processing Technology* 169:427-436.
- [18] W.S. Zhao; Q.G. Meng and Z.L.Wang. 2002. The application of research on powder mixed EDM in rough machining. *Journal of Material Processing Technology* 129:30-33.
- [19] M.Z. Zahiruddin; E.A.Rahim and S. Hassan. 2006. Effect of electrical parameters on the EDM performances of titanium alloy. In *Proceedings of the 1st International Conference and 7th AUN/SEED-Net Fieldwise Seminar on Manufacturing and Material Processing*, ISBN 983-42876-0-7.
- [20] Q.H. Zhang; R. Du, J.H. Zhang and Q. Zhang. An investigation of ultrasonic-assisted electrical discharge machining in gas. *International Journal of Machine Tools & Manufacture* DOI: 10.1016/j.ijmachtools.2005.09.023.
- [21] Jeswani M.L. 1981. Electrical discharge machining in distilled water. *Wear* 72:81-88.
- [22] S. Tariq Jilani and P. C. Pandey. 1984. Experimental investigations into the performance of water as dielectric in EDM. *International Journal of Machine Tool Design and Research* 24:31-43.
- [23] Konig W. and Siebers F.-J. 1993. Influence of the working medium on the removal process in EDM sinking. *American Society of Mechanical Engineers, Production Engineering Division (Publication) PED* 64:649-658.
- [24] Gary F. Benedict. 1987. *Non Traditional Manufacturing Processes*. Marcel Dekker Inc. New York.
- [25] Zhao F.-L.; Lu Z.-Z.; Wang H. and Qian Z.-Q. 2005. Research on effecting mechanism of particles in powder-mixed EDM. *Dalian Ligong Daxue Xuebao/Journal of Dalian University of Technology* 45:668-671.
- [26] Yan B.H.; Lin Y.C.; Huang F.Y. and Wang C.H. 2001. Surface modification of SKD 61 during EDM with metal powder in the dielectric. *Materials Transactions* 42:2597-2604.
- [27] Y.S. Wong; L.C. Lim; Iqbal Rahuman and W.M. Tee. 1998. Near-mirror-finish phenomenon in EDM using powder-mixed dielectric. *Journal of Materials Processing Technology* 79:30–40.
- [28] Fabio N. Leao and Ian R. Pashby. 2004. A review on the use of environmentally-friendly dielectric fluids in electrical discharge machining. *Journal of Materials Processing Technology* 149:341-346.
- [29] Koenig W. and Joerres, L. 1987. Aqueous solutions of organic compounds as dielectric for EDM sinking. *CIRP Annals – Manufacturing Technology* 36:105-109.

- [30] Biing Hwa Yan; Hsien Chung Tsai and Fuang Yuan Huang. 2005. The effect in EDM of a dielectric of a urea solution in water on modifying the surface of titanium. *International Journal of Machine Tools and Manufacture* 45:194-200.
- [31] P. Pecas. and E. Henriques. 2003. Influence of silicon powder-mixed dielectric on conventional electrical discharge machining. *International Journal of Machine Tools & Manufacture* 43:1465–1471.
- [32] Tzeng Yih-fong and Chen Fu-chen. 2005. Investigation into some surface characteristics of electrical discharge machining SKD-11 using powder-suspension dielectric oil. *Materials Processing Technology* 170:385-391.
- [33] F. Klocke.; D. Lung; G. Antonoglou and D. Thomaidis. 2004. The effects of powder suspended dielectrics on the thermal influenced zone by electrodischarge machining with small discharge energies. *Journal of Materials Processing Technology* 149:191–197.

The effect of moisture on the extraction rate of palm pressed fiber residue using press machine

Zainoor Hailmee Solihin*, Ahmed Jaffar^a

Faculty of Mechanical Engineering, Universiti Teknologi MARA,
40450 Shah Alam, Selangor Darul Ehsan, MALAYSIA

*corresponding author: - Tel: 6012-2646482

E-mail add: zainoor8@yahoo.com

^a E-mail add: ahmedjaffar@yahoo.com tel: 603-5543 5133, fax: 603-5543 5134

Abstract

Palm oil industry has begun in Malaysia in the early 1900's. Mechanical screw press machine with the combination of digester has been widely used in the palm oil mill to separate essential oil from the palm oil fruitlet. Other than crude palm oil and palm kernel, palm oil mill also produced palm pressed fiber which consist about 5-7% oil in wet basis as residue. The objective of this study is to investigate the effect of moisture to the extraction rate through experimental design of separating the essential oil from palm pressed fiber (PPF) residue by using press machine. The method of the experiment is by developing hydraulic press machine that simulate the operating principle of palm oil mill mechanical screw press machine. Based on the mechanical screw press machine normal operation in the palm oil mill, the simulation press model using hydraulic press was constructed which the experiment will use hydraulic press for pressing purposes. The effects of the moisture to the extraction rate were investigated and the experiment data were obtained at parameter from normal operating mechanical screw press machine at constant temperature of 90°C and constant operating pressure at 95.4kN, where to establish the optimum moisture condition.

Keywords: Palm oil, palm pressed fiber, mechanical screw press machine, oil in wet basis, moisture

1.0 Introduction

Oil Palm (*Elaeis guineensis*) was first introduced to Malaysia as an ornamental plant in 1870 and has now become the cornerstone of the country's agricultural sector[1]. Malaysia is one of the largest palm oil producing country accounting around 11.5 million ton of palm oil which is about 50% of the total world production[2]. The Palm oil industry is expected to grow 4.57% annually over the next five years to 33.39 million tons by 2010. Exports are expected to grow 5.6% to reach 23.75 million tons by 2010, capturing about 40% of the global oils and fats market[3]. Other than crude palm oil (CPO) and palm kernel (PK), palm oil mills produce significant quantities of residues such as PPF (from the mesocarp), shell (from around the kernel) and empty fruit bunches (EFB).

The palm oil processing consists of sterilizer with the objective of sterilization and to soften the palm fruitlets that will encourage the detachment of fruitlets from bunches before going to the next process of pressing station. Pressing process in the palm oil mill widely used mechanical screw

press machine shown in Figure 1 that comes together with digester in order to enhance the pressing process[4]. The screw press shown is of 9ton/hr press. Under normal circumstances, it can churn 3.15ton/hr crude oil.

The residue of PPF from the pressing process still contains about 5%-7% [5]of palm oil in wet basis. The oil content in the PPF is maintaining 5%-7% due to the limitation of press machine and the percentage of broken kernel that is still relatively high due to an increase in the pressing force during the pressing process. In the palm oil mill a 7% PPF residue of oil losses is considered high.

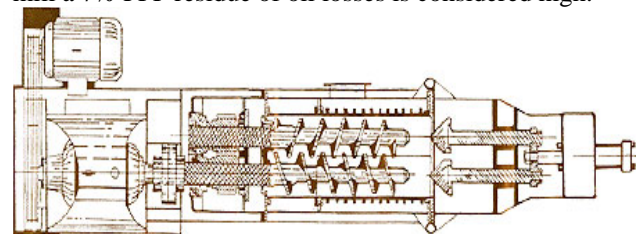


Figure 1: Mechanical screw press machine

Palm oil production generates large amounts of process residues such as fiber 5.4 million ton per year and shell at 2.3 million ton per year[6]. Press fiber amounting to 14% or 8.4 ton/h for 60ton/h mill capacity [3] and shell generated by the palm oil mills are traditionally used as solid fuels for steam boilers. The steam generated is used to run turbines for electricity production. Although much work has been done to date, more studies need to be conducted to ascertain the recovery of essential oil from PPF in order to enhance the efficiency in the palm oil mill.

2.0 Overview of Palm Oil Mill Process

The fresh fruit bunch (FFB) from the field are loaded on trucks, cages or trailer and weighed on arrival at the mill by weighbridge of 50 ton maximum capacity. The loaded FFB cages are then conveyed by the transfer carriage on the rail track and pushed into the sterilizer.

The sterilizer process shown in Figure 2 is done in 5, 7 and today 10 tons capacity FFB cages which are pushed into long cylindrical steel vessel with special doors and subjected to steam at approximately 3 bars. Majority of mills today has programmable automatic control systems to cater for proper sterilization of 90-minute cycle time at 140°C temperature[7]. After the sterilization the cages are then winched out from the sterilizer vessel by the arrangement of bollard & winch and then placed in position for the remote control overhead hoist, for the activity of emptying the FFB into the threshing machine which will separate the empty bunches from fruit.

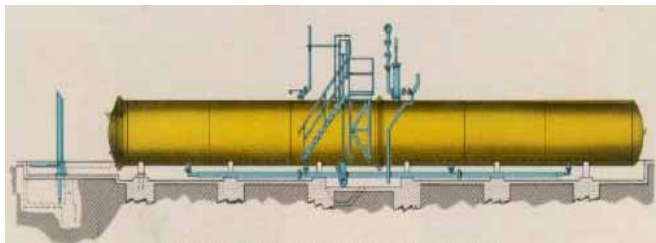


Figure 2: Sterilizer

The fruit is then conveyed by screw conveyors and bucket elevators to the pressing or extraction station. The efficient extraction of the crude oil from Tenera fruit has presented problems but these have been overcome by the development of the continuous mechanical screw press that shown in Figure 1, which is now used in all modern factories. The fruit from the stripper passes to digesters, which complete the breaking of the oil cells with slow moving arms. Digesters have a capacity of above 3 cubic meters[8]. The fruit mash then passes to the screw presses (capacities of 10–16Mt FFB per hour) which presses the crude oil out through holes in the side of the press cage[9].

The 'press cake', which is discharged from the end of the press, contains the 'fiber' and the 'nuts'. Fiber with

approximately 15% of the FFB weight has the moisture content of 37%. The residual oil content should be between 6% and 8% of oil to dry fiber [10, 11]. The fiber and nut are separated in the fiber cyclone winnowing system and the waste fiber then conveyed to the boiler as fuel as in Diagram 1. Diagram 2 shows proposed location of secondary press machine for this study where the recovered oil will be mixed with crude oil for the next process in the clarifier station[12].

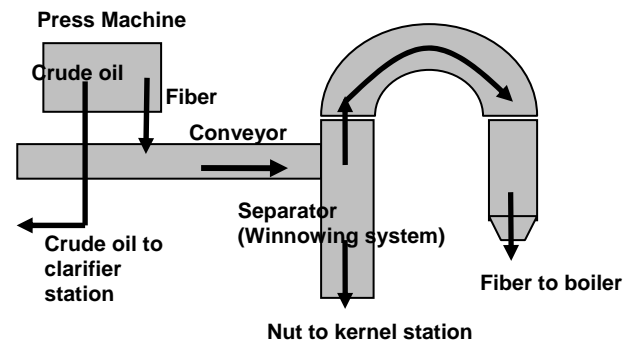


Diagram 1: Process flow in the palm oil mill, PPF and nuts from the press machine separated in the winnowing system before PPF conveyed to the boiler as fuel.

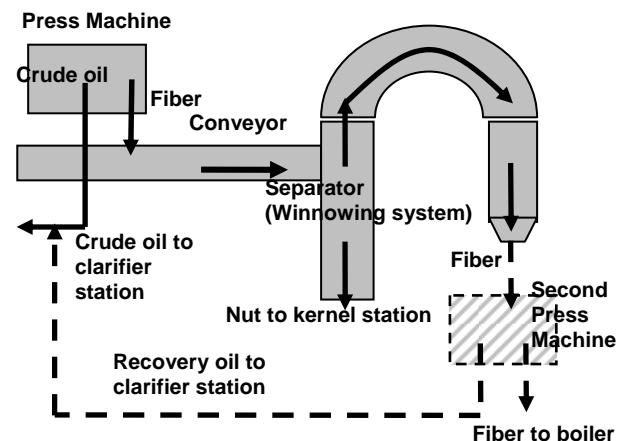


Diagram 2: Location proposal for second press machine in the study

3.0 Experimental Set up

The experimental set up for this study is shown in the Diagram 3. The experimental process starts from collecting PPF sample from the production line and ended when the waste fiber is fed as a boiler fuel. The PPF will be weight and analyze for oil and moisture content before the water is added. The water added will depend upon the moisture content in the PPF and also the weight of PPF.

Then the PPF is analyze again to determine the moisture content after the water is added. PPF will be heated in the oven to increase the temperature. Then the PPF will be pressed using the hydraulic press to extract the essential oil from PPF. Result from the extraction that are waste fiber and the mixture of water and oil will be analyzed again to determine the water and oil content.

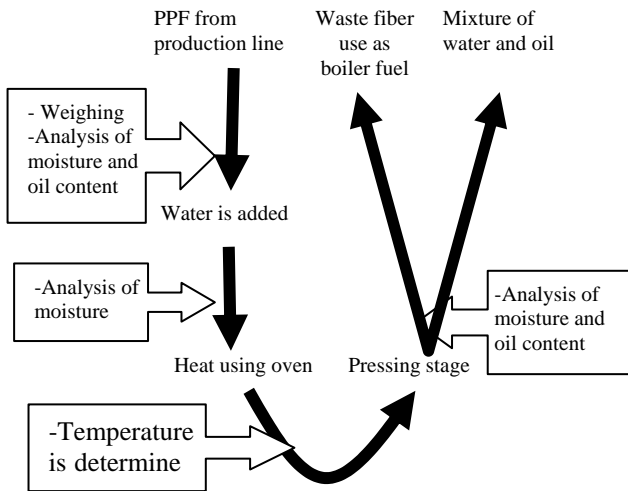


Diagram 3: Schematic Diagram of experimental flow

3.1 Apparatus

The apparatus for extraction press machine for experimental purpose is shown in Figure 3 mainly consist of hydraulic press, hydraulic pump, press cage and steam pot. The hydraulic press is fabricated using recycle material from old stork press machine and modified to suite the experimental requirement. The hydraulic press has a maximum pressing capacity of 39 ton hydraulic pump pressure as in Figure 4. The pressure can be controlled using solenoid valve and monitored from the pressure gauge.

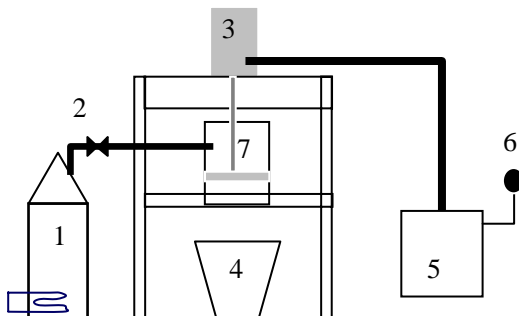


Figure 31: Diagram of experimental press machine, 1-steam pot, 2-control valve, 3-hydraulic press, 4-reservoir, 5-hydraulic pump, 6-pressure gauge, 7-press cage



Figure 4: Hydraulic pump

The weight of the press sub-assembly is about 50 kg. This part is placed at the top of the main structure. It is supported by the structure itself and fixed with bolt and nuts. It can be easily removed if necessary. The maximum recommended pressure applied is 100bars.



Figure 5: Hydraulic press

The hydraulic press structure is made of steel with dimensions of 1.84 m height and 0.91 m width as in Figure 5. It consists of the base support and adjustable work table. The work table maximum and minimum height can be adjusted from 1.60m and 0.30m respectively. The press cage has a capacity of 1.5kg to accommodate PPF sample with hole diameter of 4mm at the bottom. Steam pot is fabricated using a kettle and modified to suite the experimental requirement where the heater is to heat the water and to supply steam that is used to maintain temperature of 90°C and to obtain moisture for the PPF accordingly. The flow rate and the quantity of the steam can be controlled using the valve provided.

3.2 Material

The PPF material was directly taken from palm oil mill production line at Sepang, Selangor. PPF contained 31-34% of moisture and 5%-7% of oil on wet basis. The PPF strand has a length of about 20mm and 1-3mm wide. Oil contained from PPF was determined using Soxtech machine and the result from Soxtech machine shown the 100% of oil being extracted. The place to take the fiber is shown in Figure 6 that is after the fiber cyclone and fresh from process line to ensure the moisture, oil content and fiber quality are in good condition. Liquid use to develop moisture is water and heated using oven and the wetness of PPF is maintained using steam pot.



Figure 6: Fiber cyclone, fiber after the separation process use as a boiler fuel

3.3 Method

1 kg of sample is taken and placed in a press cage of the press machine. To study the influence of moisture on the extraction rate using press machine, a series of experiments were performed at constant temperature at 90°C and hydraulic pressure at 94kN. The particle size of PPF is ranging from 20mm long and 1-3mm wide. Water is added

to the PPF sample to increase the moisture content in PPF.

Initially sample is analyzed for the moisture content, for example the moisture is 33% and the weight is 1 kg, it takes about 70g of water to increase the moisture to 40%. The high moisture of sample will be stirred in order for water to be absorbed evenly. The sample will be heated in the oven for about 15 minutes to increase the temperature to 90°C. A small quantity of PPF sample before and after the pressing process is taken and analyzed using Soxtech machine. The result is as shown in Table 1 and 2.

4.0 Result and Discussion

The result from the experiment shows some effect on the extraction rate using press machine. Table 1 shows the results of the moisture and oil content data in the PPF before and after the pressing stage. There is no moisture added in moisture content data number 1 in Table 1. The PPF taken fresh from the production line is directly being analyzed and heated before pressing using the hydraulic press. From Table 1 there was no mixture and oil collected as the force uses in this experiment to press the PPF were identified insufficient to squeeze and collect the mixture of oil and water. Data number 1 in Table 2 show data without oil and moisture content because of no mixture can be extracted.

Data number 2 in Table 1 show some reduction of oil content in the PPF after the pressing stage and the mixture collected oil and water mixture is less as compared to data number 7 which is more diluted with water. Increment of moisture in the Table 1 shows the oil content in the PPF after the pressing has high reduction as in data number 2 and became flattered with the increment of moisture. Although the data number 2 in Table 1 has low moisture but it contains more oil compared to others as shown in Table 2.

Figure 7 show the different of oil and moisture content taken before and after the pressing stage. The graph shown at data number 2 the oil start to reduce and increase at high curve but became flat as the moisture content difference is larger before and after the pressing stage.

Table 1: Experimental data on moisture and oil content in wet basis.

Item	Average data of PPF before extraction			Average data of PPF after extraction		Different of moisture content before and after pressing (%)	Different of oil content before and after pressing (%)
	Moisture content (%)	Oil content (%)	Moisture content after add water (%)	Moisture content (%)	Oil content (%)		
1	33.51	5.57	33.51	33.51	5.57	0.00	0.00
2	31.14	6.34	43.04	33.51	4.19	9.89	2.15
3	31.52	6.24	48.84	32.30	3.85	16.54	2.39
4	34.32	5.58	54.77	33.41	3.68	21.36	1.90
5	32.32	6.03	56.34	31.23	3.62	25.11	2.41
6	33.65	6.12	59.01	33.31	3.05	25.70	3.07
7	31.56	5.92	60.22	34.12	3.11	26.10	2.81

The oil data content in Table 1 is taken after the pressing stage and the oil content in the fiber reduces proportionally with the increment of moisture content. After a certain period of moisture increment, the oil content in the PPF becomes less. This is due to limitation of water to be absorbed in the PPF. Increment of water only increases and over dilutes the PPF and the final mixture has more water content as in the result from Figure 8 but not reducing significantly the oil content in the PPF after the extraction.

Table 2: Average data on mixture collected from the PPF extraction

No.	Oil in wet basis (%)	Moisture (%)
1	0	0
2	0.87	89.54
3	0.67	91.25
4	0.61	94.33
5	0.42	94.63
6	0.33	96.22
7	0.28	97.81

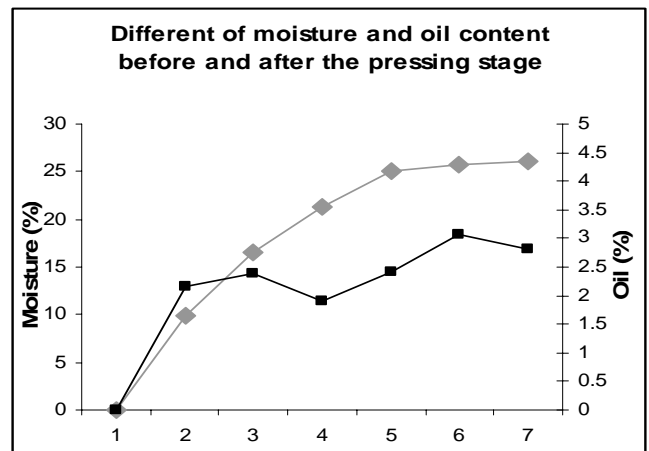


Figure 7: Different of moisture vs. .different of oil content in wet basis before and after the pressing stage

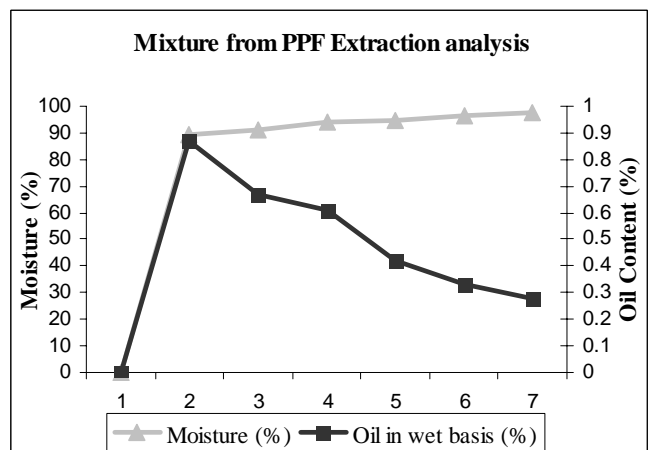


Figure 8: Moisture and. oil content in the mixture collected from the PPF extraction

5.0 Limitation

There will be some factor affecting the result in determining the moisture data level collected. Each weight of PPF per kg for the pressing experiment will be taken up to ten samples for the moisture analysis. This is because the moisture level varies depending on the water being equally absorbed in the PPF and more samples tested will bring more accurate data. The efficiency of the hydraulic press machine also affect the experimental data. In this study proper maintenance is done to ensure the hydraulic press can work efficiently.

The quantity of mixtures result from the pressing is not being quantified. This is because the most important is the oil content in the mixture. High level of oil content can save water usage and in the palm oil mill it is easier for the oil to be settled down in the clarifier station.

6.0 Conclusion

Through the extraction experiments, the effects of moisture to the extraction rate of PPF residue using press machine were studied respectively.

The extraction of PPF ranging from 40% to 50% of moisture is sufficient for the optimal moisture condition shown in Figure 7. This is because higher increment of water will only over dilutes the PPF and gives much similar effect to the oil content in the PPF after the extraction. High oil content from the mixture of oil and water collected is more preferable compared to high moisture in the mixture.

7.0 References

- [1] T.P.PANTZARIS, *Pocketbook of Palm Oil Uses*, Fifth Edition ed: Malaysian Palm Oil Board (MPOB), Sept 2000.
- [2] Z. A. Z. Z.Husain *, M.Z. Abdullah, "Analysis of biomass-residue-based cogeneration system in palm oil mills," *Biomass and Bioenergy*, vol. 24, pp. 117 - 124, 2003.
- [3] S. Yusof*, "Renewable energy from palm oil - innovation on effective utilization of waste," *Journal of Cleaner Production*, vol. 14, pp. 87-93, 2006.
- [4] M. O. F. Oseni K. Owolarafe, Obafemi O. Ajibola, "Comparative evaluation of the digester-screw press and a hand-operated hydraulic press for palm fruit processing," *journal of Food Engineering*, vol. 52, pp. 249-255, 2002.
- [5] *Mogana Report*, vol. 1: Palm Oil Research Institute of Malaysia, 1955.
- [6] Z. Z. Z. Husain *, Z. Abdullah, "Briquetting of palm fibre and shell from the processing of palm nuts to palm oil," *Biomass and Bioenergy*, vol. 22, pp. 505 - 509, 2002.
- [7] S. P. a. P. PRASERTSAN, "BIOMASS RESIDUES FROM PALM OIL MILLS IN THAILAND: AN OVERVIEW ON QUANTITY AND POTENTIAL USAGE," *Biomass and Bioenergy*, vol. 11, pp. 387-395, 1996.
- [8] G. S. C. s. A. N. V. Asterdam, *STORK Palm oil Review*, vol. 1, 1 ed. Amsterdam: GEBR.Stork & Co's, 1960.
- [9] N. R. Menon, "Evolutionary Changes In Milling Technology," in *Palm Oil Engineering Bulletin*, 2002.
- [10] N. Wambeck, *Oil Palm Process Synopsis*, vol. 1, 2 ed, 1999.
- [11] D. Y. b. b. Dato' Dr Hj Abdul Halim b. Hj Hassan, Prof.dr. Jalani b. Sukaimi, Dr. Ioke Kwong hung, Hussin b. Salleh, Mohd Nasir b. Hasan Basri, *Palm oil factory process handbook (General Description of the palm oil milling process)*, 4 ed: Palm Oil Research Institute of Malaysia, 1987.
- [12] *Mogana Report*, vol. 2: Malaysian Palm Oil Board, 1955.

A Modeling and Simulation of Process Improvement in Polyurethane Injection Manufacturing Line

Noriah Yusoff¹, Ahmed Jaffar²

^{1,2} Faculty of Mechanical Engineering
University Technology MARA, 40450 Shah Alam, Selangor, Malaysia

¹norieyusof@yahoo.com

Tel: 012-6546554

²ahmedjaffar@salam.uitm.edu.my

Tel: 03-55435155, 03-55435159

Abstracts

Manufacturing industries have been trying to reduce production time and cost of their products in various ways such as robotics applications and line automation. In today design driven and customer-centric businesses, products advancement and personalized designed is progressing very rapidly. Simulation modeling or virtual manufacturing simulation is an approach to achieve such goal. A case study was conducted at polyurethane injection manufacturing line, to model and simulate the process improvement in a car seat polyurethane (p/u) injection line using Delmia QUEST[®] simulation software. The case study subject is in a small and medium scale industry category in Malaysia and located in Klang Valley. This paper intends to discuss how Delmia QUEST[®] simulation software is used in the process improvement activities by incorporating robotics application and line automation in the simulation model. The process improvement line was modeled base on the actual manufacturing activities, using the benchmark data which was obtained during the in-field study by means of work study method. The work study method was exploited very extensively during the in-field study. The simulation model was validated by comparing the simulation results with the actual production operational data of the car seat polyurethane injection line. Financial analysis was then carried out in determining the Payback Period to justify the actual investment cost before it is materialized. A cost comparison between proposed methods against the existing method was done to put forward the improvements in terms of productivity and yearly revenue. Eventually cost effectiveness achieved; through layout improvements, line automation and robotics applications.

Keywords: Manufacturing Simulation, Line Automation, Robotics Application, Delmia QUEST[®] Simulation Software.

1. Introduction

The development of process technology and development of new improved products is important in the process industries with the aim to be cost-competitive [1]. Flexibility in production line is vital because product life cycles are getting shorter, lot size is getting smaller and there are many variants. Capability flexibility and capacity flexibility refers to the ability of the production line to react to the market demand changes such as product variants and ordering quantities, respectively [4]. Therefore, in meeting those challenges robotics and automation is of great importance. According to Groover et al, robotics and automation are closely related technologies, automation as a technology that is concerned with use of mechanical, electronics, and computer-based systems in any operations and control of productions [7]. In another explanation, Cohen and Apte has expressed a similar view that, robotics is one of the supporting elements in Production Automation Technologies apart from numerical control (NC), CNC machines and automated assembly line [8]. An automated assembly system relates to mechanization and automation of tasks in an assembly line

and robots have been commonly used in the automated assembly systems [5].

In another perspective of applying the robotics and automation technology, Cusack reveals that, the possibility of applying this technology in construction area is very promising; however when considering applying the robotics and automation technology in the construction area, the design stage of any new building is the most crucial part [9]. Whereas, Biros reviewed about the various features of using integrated robotic system in pharmaceutical company [10]. However, automotive industries still the major segment of industrial robotics and automation application. According to Weimer, in North America, automotive manufacturers were the first to apply robotics and until today they are still the major users [11]. Kusuda added to the statement by reporting, in Japan the automotive industry is still the major segment of the industrial robot application mainly in the arc welding and spot welding operation [12].

Simulation has been used extensively in designing and analyzing manufacturing systems. Dr. William A. Gruver, professor of Engineering Science and Head of IRMS

Laboratory stated that, Quest Simulation Software is not only enables them to test the validity of assumptions and evaluate the effectiveness of new algorithms, it also provides a means of combining physical and virtual systems in an operational environment [6]. This testimony was further confirmed by Bodner and McGinnis who reviewed that, simulation is a powerful tool to analyze manufacturing systems for the purpose of designing the operation or improving the present operation [13]. With the recent simulation modeling versatility, it has enabled development of any disciplines of manufacturing systems. In a separate study about the application of simulation program in manufacturing improvement activity, Lawrence affirmed that Discrete Event Simulation is primarily seen as a tool for analyzing and hence solving complex transaction processing problems [14]. It has also a significant role to play in industry and its role is further enhanced by the inclusion of animation in the study.

2. Approach and Methods

2.1 The In-field Studies: The Case Study Overview

The case study subject is in a small and medium scale industries category in Malaysia. It is a rapidly growing p/u based company that molding flexible foams for automotive applications such as front and rear seats (car seats) and floor pan insulation (for sound insulation or acoustic purposes). In this research studies, the scope of the studies only focused on the car seats products and its production line. Originally, the company had produced three car seats models namely, Spectra, Sonata and Elantra. Due to increased in demand and its present manual production line set-up, only the Spectra model can be supplied to its client instead of all the three models. In order to meet the rising demand of its client, the company has to improve and develop the present production line. This has driven the company to look seriously at its production capacity with particular emphasize on the process improvement such as line automation and industrial robotics application.

The main objectives of the process improvement are to eliminate the inherent waste or non-value adding activities in the present process and reducing lead time. Thus, more consistent quality and cost effective products can be achieved through well organized process of line automation and industrial robotics applications. By using QUEST simulation Software, this aim can be virtually visualized and optimized before the plant or production line may even exist. According to Brown, by simulating the actual product that will be produced, in any manufacturing environment, countless mistakes are avoided and many improvements can be made [15]. A related objective was to exploit is a work study methods and techniques in obtaining the case study data throughout the on-site study activities. Abdul Rahman Abdul Rahim and Mohd. Shariff Nabi Baksh affirmed that, there are no specific rules to follow when designing and conducting

case study research, though there are often similarities in approach [2]. Nevertheless, some of the tools and techniques offered by the work study methods; in particular, method study; can serve the organization in measuring progress in terms of reduced lead-time, lower cost and faster responses to delivery, which are of important factors in providing customer satisfaction and therefore, quality [3].

2.2 Layout of the Car Seat P/u Injection Line:

Figure 1: The Present Layout and the Car Seat Products (1a, b, c, d& e)

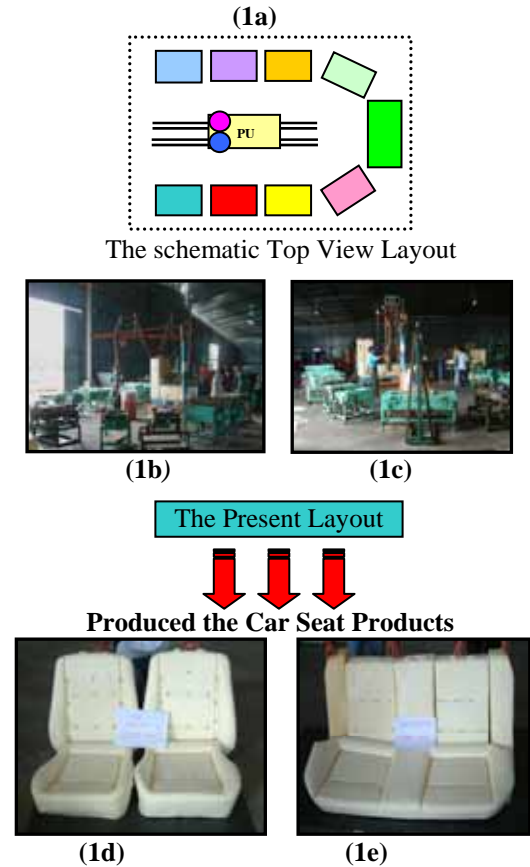


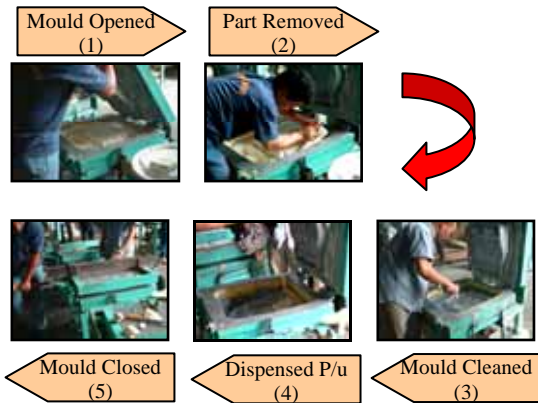
Table 1: Abbreviations of Moulds Position

1	Front Seat Cushion R	6	Rear Seat R
2	Front Seat Cushion L	7	Rear Seat L
3	Front Seat Back R	8	Bolster R
4	Front Seat Back L	9	Bolster L
5	Rear Seat Cushion		P/u machine

Figure 1 shows the present layout of the car seat p/u injection process and the car seat products. The layout area is 26 meter length by 13 meter width. The line was set-up to run three different models, which are Sonata, Elantra and Spectra. Each model comprises nine moulds to form one complete set of car seat (one car seat unit). The production has to setup the

line every time the model changes. However, increased in demand forced the production operation to concentrate only on one model, which is the spectra model. Nine moulds of Sonata and Elantra respectively, were transferred out to another sub-contractor in order to fulfill the customer’s demand. Figure 2 is a p/u injection process flow, highlighted the present method of the car seat p/u injection line

Figure 2: The Present Method (Process Flow)



2.2 QUEST® Simulation Features

QUEST® is a flexible object based discrete-event simulation tool to efficiently model, experiment and analyze facility layout and process flow. It provides 3D digital factory layout visualization and simulation analysis of the designed production control. The new improved models of the p/u injection for car seat line were constructed using the QUEST® based on the concept of the present production setup. The p/u injection process parameters such as, p/u dispensing time, machine cycle time (to represent the molding time), agv speed (to represent the robot’s movement), and conveyor speed were input into the simulation model. These inputs were obtained from the data collate during the on-site studies using work measurement - time study technique.

2.3 Assumptions and Devices

In most manufacturing systems there is a considerable amount of human involvement, this radically increase the unpredictable of the system. To account for these extreme situations, a list of assumptions was compiled. Some of the assumptions are listed below.

- **AGV(Automated Guided Vehicle); Robot**
An AGV is a Material Handling System (MHS) element. It is a predefined material handling construct that can transport multiple parts from one AGV decision point to another. Since QUEST® did not aid robot’s kinematics simulation, the AGV is to represent the movement of the robot. In the

conceptual model, the robot is stationed at one designated position and its task is to pour the p/u compound into each mould using the p/u injection machine.

- **AGV Controller; Control panel at the p/u machine.**
An AGV controller is an MHS element. It is used to widely control one or many AGV classes with respect to a set of AGV decision point. In this simulation the AGV controller denote the role of a robot’s controller is to command every movement of the robot tasks.
- **Source; P/u (Polyurethane) compound in the liquid form and stored in a tank.**
Source is an element that produces parts. In this simulation the source represents the p/u compound reservoir that supplied to the AGV (robot) to be distributed to each moulds. The source input represents the p/u dispensing time.
- **Parts; Liquid Polyurethane.**
A part is an entity that is created dynamically as the simulation runs. It moves between elements and is processed by the system. Parts are usually generated at the source or as a result of a machine process. In this simulation model the parts will be assumed in the form of solid parts instead of liquid and will be distributed by the AGV to each mould accordingly.
- **Machine; Mould**
A machine is an element that behaves much as a physical machine does. The machines symbolize the moulds in the actual process. The machine cycle time input data is actually the molding process cycle time and parts curing time inside the mould.
- **Sink; Temporary Store**
Sink is an element that used to store temporarily the molded parts before the next process. The sink input time represents the parts removing time from the mold.

2.4 Input Data and Benchmark Parameters

The input data for the simulation model was obtained from the in-field study that was conducted at the car seat line.

Table 2: Key Input Parameters for the QUEST® Model and Their Distributions

Variables	Distributions
Machine cycle time	Exponential*
Source	Poisson**
Sink	Poisson**
AGV speed	Constant
Conveyor speed	Constant

* Used to model random arrival, breakdown process, inappropriate delay times, situation where the random quantities satisfy the memory less property. E.g. Machine operational study such as machine breakdown and manpower problem.

**Used to model the number of random events occurring in an interval of time and to model random variations in batch size. E.g. Manufacturing output study such as manufacturing defects and quality control activities

The actual production output data was extracted from the Production Department records and the simulation input data was accomplished from the time study measured during the in-field studies. Table 2 indicates some of the key parameters of the simulation and their associated probability distributions.

Table 3. Benchmark Parameters for the QUEST® Simulation Modeling System

Variables	Distribution (mean) - seconds	
	100% efficiency	95% efficiency
Source	26.39	27.7
Sink	35.21	36.97
Machine	60.9	63.95

Table 3 shows the benchmark parameters that were input during the simulation modeling. The 95 % production efficiency was adopted simulating the actual production operation which is presently running at 95 % efficiency.

2.5 Constructions of the New Improved Models Using QUEST®:

Figure 5 illustrates the single work cell of dial type model with the mould layout in a circular arrangement. Figure 5 is a QUEST® simulation model of dial type with a single modular work cell. The model comprises three operators attending to the nine moulds and additional of industrial robot and conveyors were created to simulate the line automation and industrial robotics application.



Figure 5. QUEST® Model Dial Type (Single Work cell)

Figure 6 demonstrates the dial type QUEST® simulation model with three work cells running concurrently. It is an assemblage of three single modular work cells, with the same facility and setup to run three different models in tandem; Sonata, Elantra and Spectra. This would be the ultimate model that will be proposed to the case study subject. For the purpose of simulation model validation against the actual production line setup, the single work cell of dial type model simulation result will be considered, since the model is actually emulating the present production line setup.



Figure 6. QUEST® Model Dial Type (Three Work cells)

3. Results and Discussion

Simulation was done on a single modular work cell model to portray the new improved model based on the present setup condition. A final walk through was carried out to assess the model logic and ensure it was done correctly. Simulation was done for seven hours to imitate the actual production time in one shift operation. The production throughput of the simulated model was then compared against the actual production record for data analysis purposes.

3.1 Simulation results

The results of the study are shown in Table 4. The actual production output data was obtained from the production record of the case study location.

Table 4: Present Output versus the QUEST® Model

Production Efficiency	Production output (parts/day)	
	Present	QUEST® Model
100%	252	419
95%	239	409

At present, the daily production target is 252 parts/day or 28 units/day and, based on the production records, 5% rejection will be built in. Therefore, the production throughput is at 5% rejection or at 95% production efficiency and

contributing to 239 parts/day or 26.6 units/day. The simulation results indicated that the new improved model simulation mean throughput results is 419 parts/day or 46.56 units/day at 100% production efficiency and 409 parts/day or 45.44 units/day at 95% production efficiency. From this rationalization the 95 % production efficiency will be adopted in the following calculations and all the calculation will be based on one shift production schedule.

Table 5. Layout Setup of the Present Line Setup and QUEST® Model.

Items	Present Setup	QUEST® Model
Output/day	26.6 units/day	45.44 units/day
Working hrs/day	7 hrs	7 hrs
No. of workers	5 operators	3 operators
Line automation	nil	yes
Industrial robot	nil	yes

Table 6. Cycle time Comparison between Present Setup and QUEST® Model.

Line Setup	Present	QUEST® Model
Cycle time at 100% efficiency	15.00 min/unit	9.02 min/unit*
Cycle time at 95% efficiency	15.81 min/unit	9.24 min/unit*

* Cycle times for the model were obtained from the QUEST® Simulation Model.

3.2 Cost Comparison

The cost comparison was done to demonstrate the cost effectiveness and its differences between the present and new improved method. The cost comparison as shown in Table 7 explains that with the new improved model, profit per year has accelerated tremendously by approximately RM 2.4 million which is 65% increase from the present setup during its first five years of operation. Soon after five years where the new investment cost has already paid off, the profit has further stepped up to approximately RM 2.5 million which is 67% amplified from the present setup.

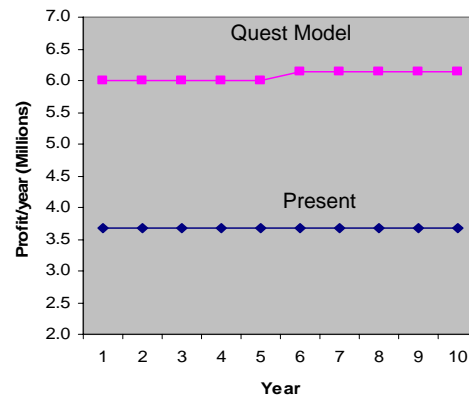
Table 7. Cost Comparison between Present Setup and QUEST® Model.

Annual Cost	Present (RM)	QUEST® Model (RM)
Profit/year	3,670,037.81	6,066,817.00
Production cost/year (8400 unit/yr)	9,175,094.53	9,107,142.40

Graph of the profit/year is illustrated in Figure 8. The production cost per unit also has reduced by 0.74% in the first five years and further reduced to 1.24 % after year 5.

Reduction in the production cost will give a better margin for the profit gained.

Figure 8. Profit/year Present versus QUEST® Model



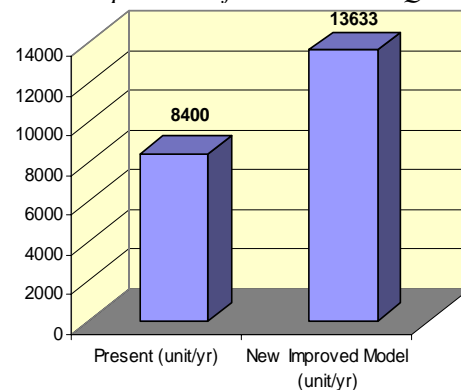
From table 8, the production volume of the QUEST® model also increased by 5233 unit/year, which is 62% higher than its present production volume. With this excess capacity gives the advantage to acquire more business opportunity. Hence, more sales can be generated; subsequently it will be reflected in the profit gained through this new improved method.

Table 8. Production Volume Comparison between Present and QUEST® Model

Volume per year	Present	QUEST® Model
Unit/year (based on line capacity)	8400	13,633

Graph of the volume per year is demonstrated in Figure 9. All the calculation above is based on one model (Spectra model) running in one shift production in a single modular work cell simulation model.

Figure 9. Volume per Year of Present versus QUEST® Model



This scenario was adopted to represent the present production condition, which is also producing one model in a single production layout setup. The production volume

achieved in the new improved model requires a production planner to schedule only two working shifts in order to fulfill their daily requirements for the three models (Sonata, Elantra and Spectra) and the extra third shift can be utilized to build up stock or expand for new business opportunity. However, if the line was setup to run all the three models concurrently, the production needs to run only in one working shift with the surplus of approximately 60% output for every models. In other words, there are huge potential business opportunity can be secured with the available capacity to offer.

4. Conclusion

Research indicates a huge potential improvement can be achieved in production volume, revenue and profit through systematic study that involves application of industrial robotics and line automation. The exploitation of QUEST® simulation software provides the means to explore into great detail study of the new improved model without interrupting the actual production line and virtually display the real life situation. Moreover, simulation provides a safe environment for manufacturers to develop new and improved manufacturing processes. From the simulation results, when the number of production unit per year increased, the sales opportunities that are available in the market can be secured accordingly. Simultaneously, when the production cost per unit price reduces, the profit per year will also increase significantly. For the benefit of the case study subject, the process improvement made on the p/u injection for the car seat line will allow the company to secure back their subcontracted model and recapture their sales revenue as well as explore other business opportunity.

5. References

- [1] Lager, T.; and Hörte, S.Å. 2002. Success factors for improvement and innovations of process technology in process industry. *Journal of Integrated Manufacturing Systems* 13(3): 158-164,
- [2] Abdul Rahman Abdul Rahim.; and Mohd. Shariff Nabi Baksh,.2003. Case study method for new product development in engineering –to-order organizations. *Journal of Work Study* 52 (1) : 25-36.
- [3] Brown, S. 1994. The role of work study in TQM. *The TQM Magazine* 6 (3): 9-15.
- [4] Heilala, J. and Voho, P. 2001. Modular reconfigurable flexible final assembly systems. *Journal of Assembly Automation* 21 (1): 20 -28.
- [5] Saad S.M., and Byrne, M.D, 1998. Comprehensive Simulation Analysis of a Flexible Hybrid Assembly System. *Journal of Integrated Manufacturing System* 9(3): 156 - 167.
- [6] QUEST discrete event simulation software to be used in systems research to evaluate and qualify best manufacturing process, October 2003, Citing Internet Sources URL http://www.tenlinks.com/NEWS/PR/delmia/101403_sfraser.htm.
- [7] Groover, M.P., Weiss, M., Nagel, R.N., and Odrey, N.G. 1986. *Industrial Robotics, Technology, Programming and Applications.*: Mc. Graw Hill, International Edition.
- [8] Cohen, M.A., and Apte U.M. 1997. *Manufacturing Automation.*: Chicago, Irwin.
- [9] Cusack, M. 1994. Automation and Robotics – The Interdependence of Design and Construction Systems *Journal of Industrial Robot* 21 (4):10 -14.
- [10] Biros, M.L. 2002. The State of HTS Automation: Comparing Integrated Robotics and Workstations. *The Int. Journal of Industrial Robotics* 29(1): 38 -42,
- [11] Weimer, G. July 2001. Robots Takes over Automotive Plant Floors. *Material Handling Mgmt Magazine*. Citing Internet sources URL <http://www.mhonline.com>
- [12] Kusuda, Y. 1999. Robotization in the Japanese Automotive Industry. *An International Journal of Industrial Robot* 26(5):358-360.
- [13] Lawrence P.J. 2003. The Multiple Role of Discrete Event Simulation in the Workplace. In 6th UICEE Annual Conference on Engineering Education, *Global Journal of Engineering Education*. 7(2): 165-172
- [14] Bodner D.A., and McGinnis L.F. 2002. A Structured Approach to Simulation Modeling of Manufacturing Systems. In *Proceedings of the 2002 Industrial Engineering Research Conference*.
- [15] Brown, J. March 2006. Digital Mnuufacturing – The Product Lifecycle Management (PLM) Approach to Better Manufacturing Processes. Tech-Clarity Inc. Citing Internet Sources <http://www.jimbrown@tech-clarity.com>.

Taguchi Methodology-based Approach to Precision grinding of Silicon.

Alao Abdur-Rasheed^{1*}, Konneh Mohamed²

^{1,2} Manufacturing and Materials Engineering Department,
International Islamic University Malaysia, 53100 Kuala Lumpur, Malaysia
Tel: +6-0126865425, Fax: +60-3-61964568, E-mail: abdur_rash@yahoo.co.uk

Abstract

Conventional grinding of hard and brittle materials such as silicon results normally in brittle fracture at the surface generating severe sub-surface damage and poor surface finish. However, recent developments in grinding technology have opened up new avenues for finishing hard and brittle materials with good surface finish. Ductile mode grinding enables material removal in brittle materials in a ductile manner rather than fracture. precision grinding of silicon in vertical surface grinding on an NC milling machine using diamond pins of diameter 5mm to produce plano surfaces has been investigated. Taguchi method was used to design and analyse the experiments. The surface roughness of ground surfaces of silicon was measured with a surface roughness tester. It was found that the surface roughness did not only depend on the controllable factors such as the in-feeds (depths of cut), feed rate and speed but also on the interactions between the speed and feed rate; feed rate and in-feeds; and among the three factors. The influence of the interactions among the controllable factors on ground surface finish has not been hitherto reported in the literature. Therefore, an attempt is made in this paper to explain further the factors affecting surface finishes on machined products.

Keywords: Vertical surface grinding, Taguchi method, Surface roughness, Ductile mode, Grinding conditions.

Introduction

Hard and brittle materials typically represented by silicon in this work are significantly important materials in electronic and optical industries but are difficult to machine. Silicon constitutes 90% of all semi-conductor materials [1, 2]. Semiconductor materials and their processing are emerging technologies of the 21st century. Semiconductors are the bases for electronic devices such as the transistors which have brought a major technological revolution in the industry and its applications have been extended to the computer industry.

The manufacture of silicon provides an edge over its competitor such as Gallium arsenide (GaAs), GaP etc. Specifically, complex electronic circuits have been incorporated on a single chip of silicon and this has revolutionised the design and manufacture of countless products [3, 4]. Silicon also finds applications in MEMS (Micro Electro-Mechanical Systems) because of its structural strength and easy miniaturization; in making* integrated circuit (IC) chips and optical components in high-resolution thermal imaging systems [5, 6]. Precision machining of silicon is needed to obtain the dimensions and accuracy required for these applications.

Several attempts have been made to machine silicon at low surface roughness with little or no subsurface damage. The concept of ductile mode machining has made this possible. Ductile mode machining is the removal of work material in a ductile manner rather than in brittle manner so that extensive micro-fracture can be minimized [7]. Single point diamond turning (SPDT) technique has been used to obtain ductile surfaces [5, 8-11]. Other researchers have used a technique known as Ultra-precision grinding (UPG) by the researchers [12-15]. Several models have been proposed to explain SPDT [16] and UPG [17] respectively. Flawless ductile mode machining, free of brittle fracture, was achieved by having a machining depth of cut not greater than a critical depth and having flattened grains protruding slightly from the surface of the grinding wheel [18].

Both SPDT and UPG require machine tools with high-accuracy servomechanisms, high loop stiffness and full flood coolant and depth of cut less than 1µm must be used [18]. These machines are too expensive thereby restricting their usage to the production of optical products. Because of this, it worth conducting experiments on less expensive machine tools. Therefore, the research was conducted around a conventional NC milling machine tool (NC DMU 35M Deckel Maho) attached with an ultra-precision high speed jig grinder (NSK Planet 850).

Traditional experimental design methods have been hitherto used by many researchers. However, in this work,

* Corresponding Author. E-mail: abdur_rash@yahoo.co.uk,
Tel: +6-0126865425, Fax: +60-3-61964568

Taguchi experimental design method was used which lessened the cost and time of performing the experiments while providing essential information. The issues discussed here are the surface roughness, micro fracture, grinding conditions which are controllable factors, the interactions among them and the effects of these factors as well as those of their interactions on the surface finish of ground surfaces of silicon.

Experimental design, set-up and Procedure.

Taguchi experimental design

The experiments were designed in accordance with the Taguchi method and appropriate orthogonal array was used [19]. Taguchi method provides smaller, less costly experiments that have high rates of reproducibility. For a study involving 13 factors at 3 levels, the full factorial requires conducting (3^{13}) 1,594,323 experimental runs. This is extremely time consuming and expensive. Using fractional factorial method for the study above would require high level of statistical expertise which may not possessed by the engineers. Another method used in the design and conduct of experiments is the traditional method. This involves studying a factor at a time. Traditional method is not only time consuming and costly but also does not take into account the influence of interactions between factors thereby making experimental conclusions irreproducible and sometimes disappointing [19]. Taguchi experimental method is an effective method that can be used and it takes into account the interactions between factors while producing smaller, less costly experiments that have high rates of reproducibility. Using the appropriate orthogonal array, a study involving 13 factors at 3 levels can be conducted with only 27 experimental runs $L_{27}(3^{13})$. Therefore, Taguchi Method was used. The details of the Taguchi method can be found in reference [19].

The objective of this research is to obtain the grinding conditions that will provide a least surface roughness measured in terms of arithmetical mean deviation (R_a) from the set grinding conditions. The controllable factors are the factors whose values can be set and maintained. For this project, the control factors identified are as specified in Table 1. The noise factors are factors whose values can not be set or we do not want to set. The ideal situation for dealing with measurement noise is to have measurements taken in a controlled environment. Therefore, the experiments were conducted in a controlled environment and there was no room to include noise factors in the experimental design.

To ensure that no vital interactions among the controllable factors are left out in the experimental design for this project, all possible interactions between the factors were incorporated. However, they were limited to interactions between two controllable factors only as three-way interactions (interaction between factors A, B and C) might be insignificant according to the literature [19].

The controllable factors and the possible interactions were set at three levels. Three levels for these factors can be used to evaluate non-linearity over the experimental range [19]. Table 1 summarises the controllable factors used in this experiment and their levels while Table 2 gives the summary of their interactions.

Table 1: Controllable factors and their levels

Controllable factors	Level 1	Level 2	Level 3
A: Pressure (Spindle Speed) Kg/cm ²	2.5 (40,000 rpm)	3.0 (55,000 rpm)	3.5 (65,000 rpm)
B: Feed rate (mm/min)	2.5	5	10
E: Depth of cut (μ m)	5	10	15

Table 2: Interactions settings

Interactions between	Level 1	Level 2	Level 3
Factor A and Factor B: AB	1	2	3
Factor A and Factor E: AE	1	2	3
Factor B and Factor E: BE	1	2	3

Experimental set-up and procedure

Silicon material having the following properties shown in Table 3 was precision cut into smaller samples of 15 x 15 x 12mm. Special fixtures were designed to hold the workpiece geometry. An ultraprecision high speed jig grinder (NSK Planet 850), Figure 1, was attached to the vertical spindle of a NC milling (DMU 35M Deckel Maho) having a x-y-z axis resolution of 0.001mm (1 μ m). A rectangular pocketing was applied to the samples to remove the uneven surface after precision cutting and also to set the zero point for the depth of cut. Grinding pins (1A1W-5-6-1.5/D64/C100) also known as resinoid-bonded mounted wheels were used. Other details about the grinding pins are shown in Figure 2.

The grinding pins were ultrasonic drilled to have a central cavity in order to remove stalling marks and zero velocity tracks phenomena [20]. An Al₂O₃ abrasive stick was used to dress the pins at 10 μ m in-feeds in every pass before starting a new grinding condition. This was done in the presence of a coolant. The coolant was Castrol Miracol 80 with ratio of 1:50 based on past work [21]. The coolant was kept constant throughout the experiments. Experiments were conducted following the Taguchi method in Table 3. The surface roughness was measured using Mitutoyo Surf test (SV-514). Each experimental run was repeated and the average value of surface roughness of each experiment was noted.



Figure 1: The experimental set-up showing a high-speed jig grinder unit attached on NC vertical milling spindle and the Al₂O₃ abrasive stick (grade Nr.2).

Table 3: Some properties of silicon [3]

Lattice structure	Diamond cubic
Bond type	Covalent
Lattice constant Å	5.4307
Density g/cm ³	2.328
Melting Point °C	1410
Young modulus, N/mm ²	188.400
Poisson's ratio,	0.28
Hardness, H _v	950
Mohs Hardness	6.5
Fracture Toughness (MPa mm ^{1/2})	15
Resistivity at 300 K,	2.3*10 ⁵
Energy Gap, eV	1.1

Result and Discussion

The results of the experiments are shown in Table 4. They were analyzed following the laid down procedure by Taguchi method [19]. It is seen from the Table 4 that the quality characteristic of interest is the surface roughness, i.e. the smaller the surface roughness the better. Also seen in this table are the columns C, D, F, G, H, L which are used as interaction columns. Columns J, K, M, and N are error columns left in the orthogonal array to test if there is any interaction among the three factors that influence greatly the quality characteristic. Since each column takes two degrees of freedom (df), the error columns should have 8 df [19]. The results were therefore analysed as follows.

Step 1. Calculate the effect of each factor on the surface finish. This was done with the aid of Qualitek 4 software and the following responses were generated as shown in Table 5. L1, L3, and L3 are the effects of each factor or interaction as

set at three levels while the δ row represents the difference between the largest and the smallest values. For example, for factor A, the largest value is 0.112 and the smallest is 0.095, so δ is 0.017. For factor B, the largest value is 0.145 and the smallest value is 0.056, so δ becomes 0.089. Other δ values were calculated in the same fashion.

The factors that influence the surface finish are typed in bold faces. Factor B has the greatest effect on the surface roughness followed by factor N and so on. The difference between factors B and N is 0.042. Other calculations were done in the same way. Since the difference between factor K (error 2) and factor J (error 1) which is 0.006 is greater than the difference between factor F (AB₁) and factor K (error 2) which is 0.001, factor K was ruled the logical breaking point [19] for this work. Based on the response table, Table 5, Figure 3 which shows the influence of main effects on the surface finish of ground silicon samples is thus constructed.

Referring to Table 5 and Figure 3, since error columns J, K, M and N were used to test the effect of the possible interaction among the three factors and column N from Table 5 has a higher effect than column K, the inclusion of column N (Err 4) nullifies the inclusion of other columns J, K, M. The same reason is applicable to the nullification of columns L (BE₂) for the interaction BE; G (AE₂) for AE and D (AB₂) for AB respectively.

Step 2: Detailed study of the interactions AB₁, AE₁ and BE₁, represented by columns C, F and H respectively in Table 5 was carried out to determine the appropriate values to be included in the prediction equation. For the interaction BE in column H, the response values in the average R_a column were for column B and E at each level and the average values were calculated. For example, for B₁E₁, the R_a for row B₁ and E₁ were averaged.

$$\text{Mathematically, } B_1E_1 = (0.10 + 0.04 + 0.10)/3 = 0.08 \quad (1)$$

$$B_2E_2 = (0.13 + 0.14 + 0.28)/3 = 0.18. \quad (2)$$

Other values were calculated in the same similar way and the following response table for the interactions BE, AE and AB

- Shape: 1A1W
- Grit size/Concentration: 64µm/C100
- Diameter (D): 5mm
- Height (T): 6mm
- Diamond thickness (X): 1.5mm

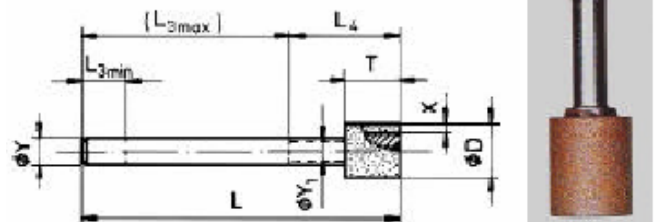


Figure 2: Details of the grinding pins used to carry out plano grinding.

Table 4: Taguchi Orthogonal Array and the result for the experiments.

A: Pressure (Kg/cm ²)	B: Feed rate (mm/min)	C: AB ₁	D: AB ₂	E: Depth of cut (μm)	F: AE ₁	G: AE ₂	H: BE ₁	J: Err 1	K: Err2	L: BE ₂	M: Err 3	N: Err 4	Response Average Ra (μm)
2.5	2.5	1	1	5	1	1	1	1	1	1	1	1	0.10
2.5	2.5	1	1	10	2	2	2	2	2	2	2	2	0.13
2.5	2.5	1	1	15	3	3	3	3	3	3	3	3	0.09
2.5	5.0	2	2	5	1	1	2	2	2	3	3	3	0.09
2.5	5.0	2	2	10	2	2	3	3	3	1	1	1	0.13
2.5	5.0	2	2	15	3	3	1	1	1	2	2	2	0.16
2.5	10	3	3	5	1	1	3	3	3	2	2	2	0.06
2.5	10	3	3	10	2	2	1	1	1	3	3	3	0.04
2.5	10	3	3	15	3	3	2	2	2	1	1	1	0.06
3.0	2.5	2	3	5	2	3	1	2	3	1	2	3	0.04
3.0	2.5	2	3	10	3	1	2	3	1	2	3	1	0.11
3.0	2.5	2	3	15	1	2	3	1	2	3	1	2	0.14
3.0	5.0	3	1	5	2	3	2	3	1	3	1	2	0.14
3.0	5.0	3	1	10	3	1	3	1	2	1	2	3	0.14
3.0	5.0	3	1	15	1	2	1	2	3	2	3	1	0.12
3.0	10	1	2	5	2	3	3	1	2	2	3	1	0.06
3.0	10	1	2	10	3	1	1	2	3	3	1	2	0.07
3.0	10	1	2	15	1	2	2	3	1	1	2	3	0.07
3.5	2.5	3	2	5	3	2	1	3	2	1	3	2	0.10
3.5	2.5	3	2	10	1	3	2	1	3	2	1	3	0.12
3.5	2.5	3	2	15	2	1	3	2	1	3	2	1	0.11
3.5	5.0	1	3	5	3	2	2	1	3	3	2	1	0.18
3.5	5.0	1	3	10	1	3	3	2	1	1	3	2	0.28
3.5	5.0	1	3	15	2	1	1	3	2	2	1	3	0.07
3.5	10	2	1	5	3	2	3	2	1	2	1	3	0.05
3.5	10	2	1	10	1	3	1	3	2	3	2	1	0.05
3.5	10	2	1	15	2	1	2	1	3	1	3	2	0.05

Table 5: Response table

	A	B	C	D	E	F	G	H	J	K	L	M	N
L1	0.095	0.104	0.116	0.096	0.091	0.114	0.088	0.083	0.109	0.117	0.107	0.097	0.102
L2	0.098	0.145	0.091	0.101	0.118	0.085	0.106	0.105	0.105	0.093	0.097	0.104	0.125
L3	0.112	0.056	0.098	0.108	0.096	0.106	0.111	0.117	0.091	0.095	0.101	0.104	0.078
δ	0.017	0.089	0.025	0.012	0.027	0.029	0.023	0.034	0.018	0.024	0.01	0.007	0.047

are shown in Tables 6, 7, 8 respectively while Figures 4, 5 and 6 show the interaction graphs for the three interactions. Since no lines are intersected in Figure 6, then there is no extremely strong interaction between the speed and feed rate compared to Figures 4 and 5. That is why in the interaction analysis both tabular and graphical method must be used [19].

Table 6: B X E Matrix

	E1	E2	E3
B1	0.08	0.12	0.113
B2	0.17	0.18	0.06
B3	0.05	0.05	0.06

Table 7: A X E Matrix

	E1	E2	E3
A1	0.083	0.10	0.103
A2	0.08	0.16	0.11
A3	0.11	0.15	0.076

Table 8: A X B matrix

	B1	B2	B3
A1	0.106	0.126	0.053
A2	0.096	0.133	0.176
A3	0.11	0.176	0.05

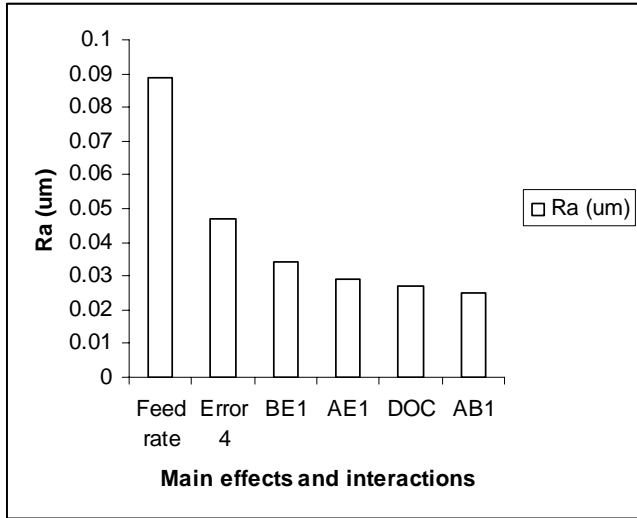


Figure 3: The effect of main effects and interactions on the surface finish of ground silicon surfaces.

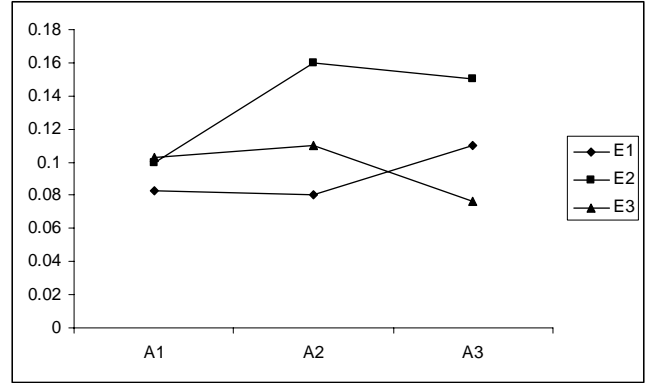


Figure 5: The graph of A X E interaction

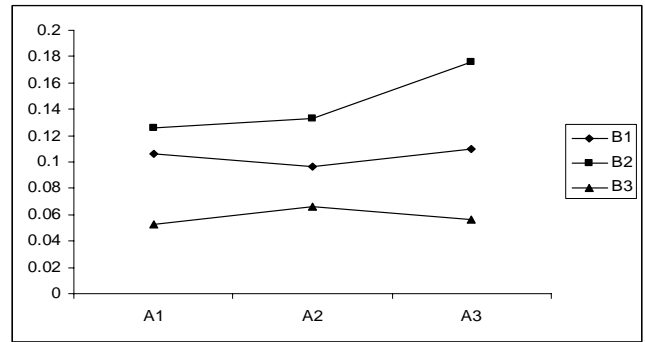


Figure 6: The graph of A X B interaction.

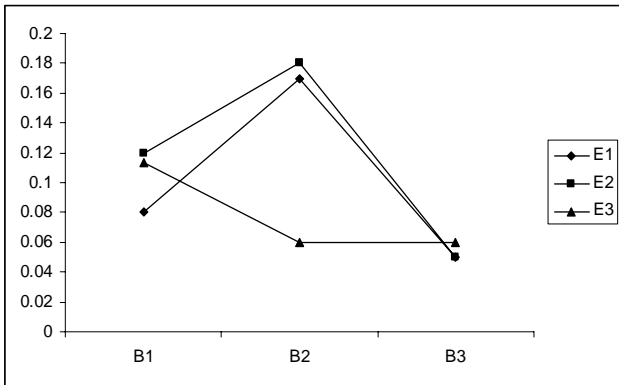


Figure 4: The graph of B X E interaction

Step 3: Setting the recommended levels for the factors and the interactions that would give the smallest surface roughness from the grinding conditions set.

The factors that have great influence on this quality characteristic are factors B (feed rate), N (Error 4), H (BE₁), F (AE₁), and E (depth of cut) and their recommended levels are B₃ (0.056), N₃ (0.078), from Table 5; B₃E₁ (0.05), Table 6; A₃E₁ (0.11), Table 7; and E₁ (0.091), Table 5

Step 4: Compute an estimate of the prediction average equation based on the selected levels of strong effects [19]. Only those values highlighted in step 3 above were included in the prediction equation. The overall experimental average (τ) was calculated as follows

$$\tau = (\sum R_a) / 27 = (0.10 + 0.13 + \dots + 0.05) / 27 = 2.76 \quad (3)$$

The prediction average was defined by \bar{u} , the effect on the overall average by each of the strong factors and interactions; the prediction equation was calculated as thus,

$$\bar{u} = \tau + (B_3 - \tau) + (N_3 - \tau) + \{(B_3E_1 - \tau) - (B_3 - \tau) - (E_1 - \tau)\} + \{(A_3E_1 - \tau) - (A_3 - \tau) - (E_1 - \tau)\} + (E_1 - \tau) \quad (4)$$

where \bar{u} was defined as the predicted average surface roughness. By combining like terms, the equation reduced to

$$\bar{u} = N_3 + B_3E_1 - E_1 + A_3E_1 - A_3 \quad (5)$$

Substituting the values,

$$\bar{u} = 0.078 + 0.0566 - 0.091 + 0.11 - 0.112 \approx 0.04\mu\text{m} \quad (6)$$

Based on this predicted value for the surface roughness, the predicted surface roughness value should be about 40nm. Experiments were later conducted to verify this using a feed rate, B₃ (10mm/min), pressure, A₃ (3.5 Kg/cm²) and depth of

cut, E_1 (5 μ m). The average surface roughness measured of the confirmation experiments was 50nm.

Conclusions.

Grinding of silicon using grinding pins has been analysed for silicon samples using Taguchi method. The results show that not only the grinding conditions like feed rate, depth of cut, and the speed have influence on the surface finish of ground silicon surfaces but also some of their interactions. A novel interaction was found to exist among the three grinding conditions and this had the second strong effect on the silicon ground surfaces after the feed rate. It was also found that there were interactions between feed rate and depth of cut; pressure and depth of cut and these were consistent with the results found in the literature [22]. Some of the silicon ground surfaces exhibited micro-fracture at higher depth of cut confirming the ductile mode phenomenon in which material is removed with little or no subsurface damage at lower depth of cut.

However, in order to generalize the effect of these grinding conditions and their interactions on the surface roughness of hard and brittle materials, much research effort has to be done on other materials like germanium and advanced ceramics using Taguchi method and their results compared. The influence of pressure may be noticed when they are set far away from one another. Also, the influence of the grinding fluid and the type of dressing method used also needs to be investigated.

References

1. Tönshoff, H.K.; Schmieden, W.V.; Inasaki, I.; König, W.; and Spur, G. 1990. Abrasive Machining of Silicon. *Annals of the CIRP* 39(2):621-635.
2. Venkatesh, V.C.; Inasaki, I.; Toenshof, H.K.; Nagawa, T; and Marinescu, I.D. 1995. Observation on Polishing and ultraprecision machining of semiconductor substrate materials. *Annals of CIRP* 44(2):611-618.
3. Komanduri, R.; Lucca, D.A.; and Tani, Y. 1997. Technological Advances in Fine Abrasive Processes. *Annals of the CIRP*. 46(2):545-596.
4. Streejith, P.S. and Ngoi, B.K.A. 2001. New materials their machining. *The International Journal of Advanced Manufacturing Technology* 18:537-544.
5. Fang, F.Z. and Venkatesh, V.C. 1998. Diamond Cutting of Silicon with Nanometric Finish. *Annals of CIRP*. 47(1):45-49.
6. Konneh, M. 2003. Experimental Investigation on Partial Ductile Mode Grinding of Silicon. Ph.D Dissertation, Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, Johor.
7. Strenkowski, J.S. and Gary, D.H. 1990. A Technique for Predicting the Ductile Regime in Single Point Diamond Turning of Brittle Materials. *Fundamental Issues in Machining PED* 43:67-80.
8. Nakasuji, T.; Kodera, S.; Hara, S.; Matsunaga H.; Ikawa, N.; and Shimada, S. 1990. Diamond Turning of Brittle Materials for Optical Components. *Annals of CIRP*. 39(1):89-92.
9. Shibata, T.; Ono, A.; Kurihara, K.; Makino, E.; and Ideda, M. 1994. Cross-section Transmission Electron Microscope Observation of diamond Turned Single-crystal Silicon Surfaces. *Applied Physics Lett* 65:2553-2555.
10. Shibata, T.; Fujii, S.; Makino, E.; and Ideda, M. 1996. Ductile-regime Turning Mechanism of Single-crystal silicon. *Journal of International Societies for Precision Engineering and Nanotechnology* 18: 29-137.
11. Venkatesh, V.C.; Fang, F.; and Chee, W.K. 1997. On Mirror surfaces with and without polishing. *Annals of CIRP*. 46(1):505-508.
12. Bandyopadhyay, B.P.; Ohmori, H.; and Takahashi, L. 1996. Ductile Regime Finish Grinding of Ceramics with Electrolytic In-process dressing (ELID) grinding. *International Journal of Material Manufacturing Process* 11:789-801.
13. Fujihara, K.; Ohshiba, K.; Komatsu, T.; Ueno, M.; Ohmori, H.; and Bandyopadhyay, B.P. 1997. Precision Surface Grinding Characteristics of Ceramic Matrix Composites and Structural Ceramics with ELID. *Journal of Machining Science Technology* 1:81-94.
14. Bandyopadyay, B.P. and Ohmori, H. 1999. The Effect of ELID Griinding on the flexural strength of Silicon. *International Journal of Machine tools and Manufacturing* 39:839-853.
15. Zhang, B.; Yang, F.; Wnag, J.; Zhu, Z.; and Monahan, R. 2000. Stock Removal Rate and Workpiece Strength in Multi-pass Grinding of Ceramics. *Journal of Material Process Technology* 104:178-184.
16. Blackley, W.S. and Scattergood R.O. 1991. Ductile-Regime Machining Model for Diamond Turning of Brittle Materials. *The Journal of the International Societies for Precision Engineering and Nanotechnology* 13(2):95-102.
17. Bifano, T.G.; Dow, T.A.; Scattergood, R.O. 1991. Ductile Regime Grinding: A New Technology for Machining Brittle Materials. *The Journal of Engineering for Industries* 113:184-189.
18. Zhong, Z.W. 2003. Ductile or Partial Ductile Mode Machining of Brittle Materials. *The International Journal of Advanced Manufacturing Technology*: 579-585.
19. Glen, S. P. eds. 1994. Taguchi Methods – A Hands-On Approach to Quality Engineering.: Addison-Wesley.

20. Izman S. and Venkatesh V.C.. 2 March, 2004.
www.fkm.utm.my/~izman/RK_Venkatesh.pdf.
21. Venkatesh, V.C.; Izman, S.; Sharif, S.; Mon, T.T.; and Konneh, M. 2003. Ductile Streaks in Precision Grinding of Hard and Brittle Materials. *Sādhanā* 28(5): 915-924.
22. Matsuo, T.; Touge, M.; and Yamada, H. 1997. High-Precision Surface Grinding of Ceramics with Superfine Grain Diamond Cup Wheels. *Annals of CIRP* 46(1):249-252.

Evaluation of a Material Flow Stress Model Adopted in Finite Element Modeling of Metal Cutting

M. N. Tamin, S. Izman, V. C. Venkatesh, T.T. Mon

*Faculty of Mechanical Engineering,
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
E-mail: pm033011@siswa.fkm.utm.my*

Abstract

The success of finite element modeling of a metal cutting process depends primarily on the material flow stress model, the numerical method and the capability of a simulation tool. While a good choice can be reliably made on the numerical and simulation sides for metal cutting simulation a flow stress model is, in most of the case, chosen without much consideration on its reliability. One reason is that experimental procedures to obtain parameters of the flow stress model, which represents work material constitutive behaviour under cutting are difficult and costly. In this study, reliability of Johnson Cook (JC) model with the parameter values for Ti6Al4V alloy obtained from the literature is evaluated by simulating tensile test of Ti6Al4V under strain up to 3, strain rate up to $100,000 \text{ s}^{-1}$ and temperature up to $1000 \text{ }^\circ \text{C}$. Commercial finite element code ABAQUS/Explicit based on updated Lagrangian method is utilized for modeling and simulation. In the absence of material testing to perform at strain, strain rate and temperature pertaining in cutting, this alternative provides reliability check on the flow stress model and relevant parameters chosen. The simulation results show that the JC model with the predetermined parameters is capable of showing material flow consistently. This is also confirmed by additional simulation of machining of Ti6Al4V alloy.

Keywords:

Flow stress, finite element modeling, Johnson Cook model, metal cutting, Ti6Al4V

1. Introduction

Finite element modeling (FEM) has been extensively used in the analysis of metal cutting processes. Usui and Shirakashi [1] developed finite element modeling of elastic-plastic workpiece cutting with a rigid sharp tool defining node separation criterion for chip formation. Iwata et al. [2] established a rigid-plastic cutting model incorporating a ductile fracture criterion. Strenkoski et al. [3] utilized an implicit finite element code "NIKE-2D" with a fracture strain criterion. All these early models did not take into account of the effect of tool edge geometry on metal cutting simulation. Recent development used remeshing methods to simulate the effect of the tool edge geometry. Marusich [4] developed a FE cutting model based on dynamic Lagrangian formulation which was later transformed into an explicit FEM code called "Third Wave AdvantEdge". Ceretti [5] used an early version of commercial implicit code 'DEFORM-2D' based on static Lagrangian formulation. Arbitrary Lagrangian-Eulerian (ALE) method is a recent development which combines the strength of Lagrangian and Eulerian avoiding the drawbacks in the previous numerical methods. With the application of ALE-based FEM, chip separation criterion can be avoided in metal cutting simulation [6]. Other ALE analysis of metal cutting can be found in the literature [7-9].

With the amazing development of FEM, the importance of a material flow model in finite element modeling of metal cutting has been repeatedly emphasized.

To ensure the chosen finite element method capable of producing reliable outputs such as chip formation, cutting forces, temperature distributions and stress distributions, an accurate flow stress data of the workpiece is mandatory. Therefore, it is very crucial to realistically determine the parameters involved in the flow stress model. However, experimental procedures to obtain such realistic parameters are difficult and costly.

Despite rapidly growing research on FEM of metal cutting, only few have done thorough analysis of the material behavior prior to the FEM. Childs [10] studied power law and JC models by analytical calculations. The results showed large difference in flow stress in the secondary shear zone. Hamann [11] investigated three constitutive models used in FEM simulation of metal cutting. He found that three different models produced significant difference in cutting ratio and maximum tool temperature. Arrozola, cited by [12], conducted sensitivity analyses of the parameter values in JC model in metal cutting simulation using ABAQUS Explicit version 6.1 and results were compared with Third Wave AdvantEdge simulation. More critical review on material constitutive behavior of some material can be found in [13].

In most published work, the parameter values for the flow stress model were taken directly from the literature and adopted in the FE model. Generally, these parameters were, in fact obtained from conventional tensile test conducted at very low strains (mostly < 1) and strain rates ($\sim 0.001 \text{ s}^{-1}$), and at room temperature, or split Hopkinson

pressure bar test at strain rates up to 1000 s^{-1} and temperatures up to $600 \text{ }^\circ\text{C}$. Parameter values associated with strain rates and temperatures beyond these ranges are then extrapolated. In reality, the material being cut will deform at very much high strains (~ 4), strain rates ($\sim 10^6$) and temperatures ($\sim 1000 \text{ }^\circ\text{C}$) [12]. Obviously, this may cause serious errors in FEM simulation results when compared to the actual experiments. In this work, JC model and its parameters, obtained from the literature, for Ti6Al4V alloy are evaluated by tensile test simulation using ABAQUS/Explicit in order to investigate the material behaviour of Ti6Al4V. Results from tensile test simulation were compared with that from the literature. The reliability of the flow stress model is then further demonstrated by simulation of machining of Ti6Al4V with polycrystalline diamond insert. Ti6Al4V was considered for analysis of material behaviour, because it is a major material in aerospace industry and troublesome material to machine; so sound research on its material behaviour is still demanding. On the other hand polycrystalline diamond tool is selected for practical machining because it has very high thermal conductivity to withstand the heat generated during cutting as cutting temperature of titanium alloys is very high. In addition using diamond tool in machining titanium alloy, built-up edge problem can be ignored and smooth chip flow on the tool can be assumed.

2. Tensile test simulation

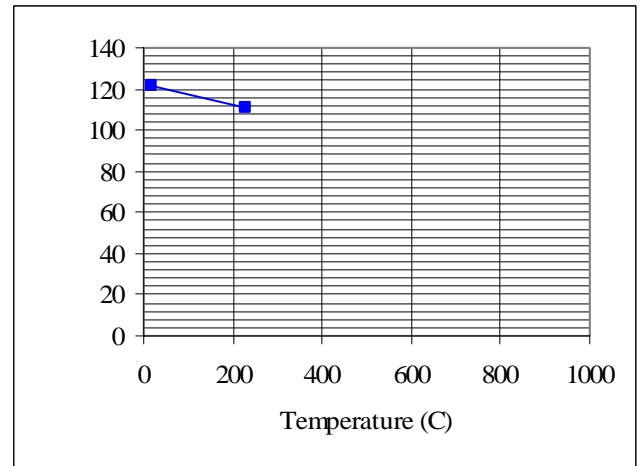
Ti6Al4V circular bar of diameter 5 mm and length 50 mm was modeled in ABAQUS/CAE. Due to symmetric nature of the bar, only a half of it was modeled to reduce computational time. The temperature-dependent properties of Ti6Al4V were obtained from TiMET, USA as shown in Figure 1. The density, Poisson’s ratio and specific heat of Ti6Al4V are 4420 kg/m^3 , 0.34 and 586 J/kg K respectively [14].

Details of the JC model were given as below:

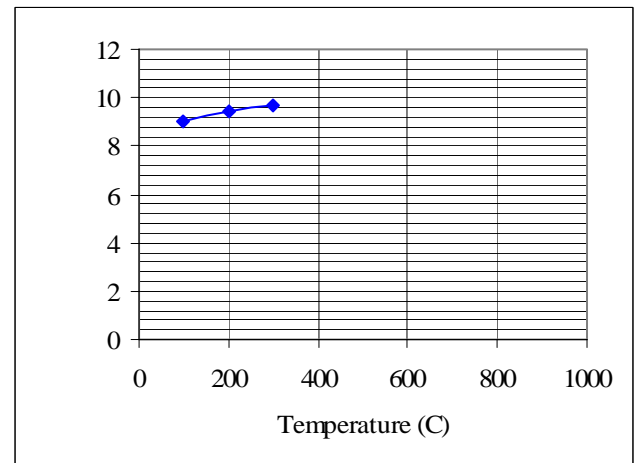
$$\sigma = \left(A + B \epsilon_{pl}^n \right) \left[1 + C \ln \left(\frac{\dot{\epsilon}}{\dot{\epsilon}_0} \right) \right] \left[1 - \left(\frac{T - 300}{T_m - 300} \right)^m \right]$$

The model describes the flow stress as a function of plastic strain (ϵ_{pl}), strain rate ($\dot{\epsilon}$), and temperature (T being transition and T_m the melting). It also considers the strain hardening effect (n) and strain rate sensitivity (m). The strain rate is normalized with the reference strain rate ($\dot{\epsilon}_0$). A, B and C represent material parameters. Generally, A, B, C, n and m are determined by several material tests conducted at low strains and strain rates, and at room temperature. JC model is chosen because it is numerically robust and fits well for strain-hardening behaviour. Many researchers has used JC model in FEM of metal cutting. Material parameters were taken from the literature as given below [13]:

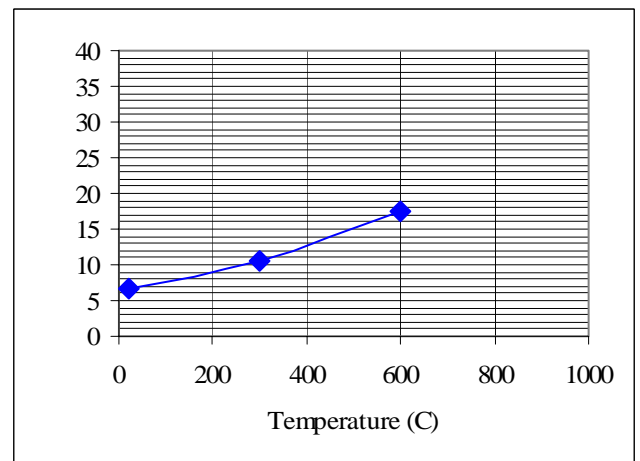
$A = 782.7$, $B = 498.4$, $n = 0.28$, $C = 0.028$ and $m = 1.0$



(a) Elasticity



(b) Thermal expansion coefficient



(c) Thermal conductivity

Figure 1. Temperature-dependent properties of Ti6Al4V [14].

Figure 2 shows the initial FE model for tensile test simulation. The bar was discretized by 8-node thermal-

displacement type brick elements available in ABAQUS/Explicit library. One end of the bar was constrained in y-direction and the displacement type boundary condition was applied to the other end for straining the bar at specified strains and strain rates. The nodal temperatures were specified to be 200, 400, 600, 800 and 1000 °C in each simulation. A thermal mechanical analysis was done with ABAQUS/Explicit solver. This analysis type solves simultaneously for mechanical as well as thermal stresses. Detailed explanation of ABAQUS/Explicit can be found in the literature [15]. A total of 10 simulations were done. Each simulation needs average 3 hour CPU time.

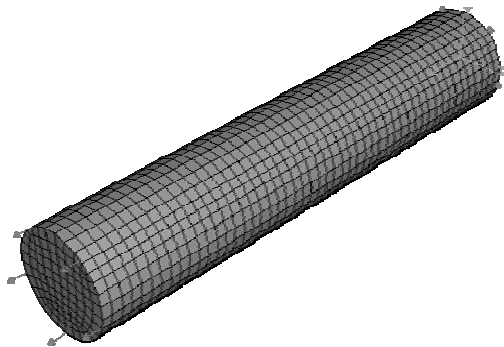


Figure 2. Finite element model for tensile test simulation.

3. Machining Simulation of Ti6Al4V

After the tensile test simulation, cutting of Ti6Al4V was modeled and simulated using the same properties input as in tensile simulation. Cutting model developed represents 2D plain orthogonal cutting which is commonly used in FE simulation of the cutting process as it provides good approximation to real cutting processes. Figure 3 shows the machining model at the initial state under plain strain condition. Both workpiece and the tool were modeled with the plain strain temperature-displacement element available in ABAQUS/Explicit element library. The properties of the cutting tool are given in Table 1. The tool's rake and clearance angles were +6 and +5 respectively. The uncut chip thickness was 0.125 mm. The bottom surface of the workpiece was fixed in X and Y directions, and the top and back surface of the tool were constrained in Y-direction only. The tool was then translated in X-direction with a specified cutting velocity of 1.5 m/s. Geometrical nonlinearity due to the shape change of finite element was taken care in the software by default setting. Heat generation at the chip-tool interface was considered due to both plastic deformation and friction.

To solve the mesh distortion problem during large deformation, updated Lagrangian formulation was adopted. Subsequently the contact conditions can be resolved correctly throughout the analysis. In order to examine the flow of the work material on the tool surface distinctly avoiding consideration of other factors such as friction at the chip-tool interface and chip separation criterion, chip formation is simulated based on pure deformation and

friction at the chip/tool interface was assumed to be zero. It should be noted that the intent here is only to demonstrate the proper flow of the titanium alloy on the tool rake face. So the detailed analysis of chip formation was not discussed here. In order to be consistent with tensile test analysis, fully coupled temperature-displacement analysis was done with ABAQUS/Explicit solver.

Table 1. Properties of PCD tool.

Properties	values
Density (kg/m^3)	3520
Young Modulus (GPa)	1200
Poisson ratio	0.2
Thermal expansion coefficient (/K)	1.5
Thermal conductivity (W/m-K)	1800
Specific heat (J/kg-K)	502

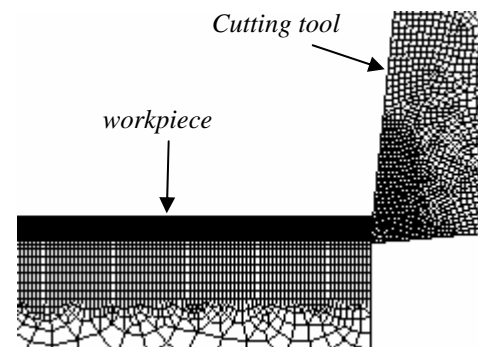


Figure 3. Machining model.

4. Experimental validation procedure

Turning of Ti6Al4V material using polycrystalline (PCD) insert was carried out to compare FEM results and the experiments. Experiments were conducted on MAHO CNC turning center using the same cutting conditions as in FEM. Ti6Al4V cylindrical bars are supplied by TiMET, USA. To achieve orthogonal cutting operation with required rake and clearance angles, PCD insert of triangular shape (ISO designation: TPGN 160308F) was used together with CTGPL tool holder. The holder and inserts were obtained from Kennametal. This insert-holder combination gives the rake angle +6 and the clearance angle +5. Width of cut was set to be 1.25 mm. Since the width of cut is much larger than the feed (0.125 mm), plain strain condition is satisfied. Water-based fluid (5% Castrol Miracol 80) was supplied to the cutting vicinity during turning to aid reducing cutting temperature and thus to protect early thermal damage to the tool. The turning experiments were repeated twice with these conditions. The inserts were then investigated under Scanning Electron Microscope.

5. Results and Discussions

The stress contour propagations in tensile test simulation were not significantly influenced by the change of strain rate and temperature. Therefore contour plot for one set of condition was shown as an example. Figure 4 shows the contour propagation at strain rate 10 s^{-1} and temperature $400 \text{ }^\circ\text{C}$. In all cases, necking where maximum stress is imposed occurred at the central portion of the bar. The location of the necking indicates that the FE model of tensile test well predicts the proper deformation of the bar. The stress contours also show a reasonable distribution of stress throughout the bar propagating gradually from a minimum at the fixed end to a maximum at necking.

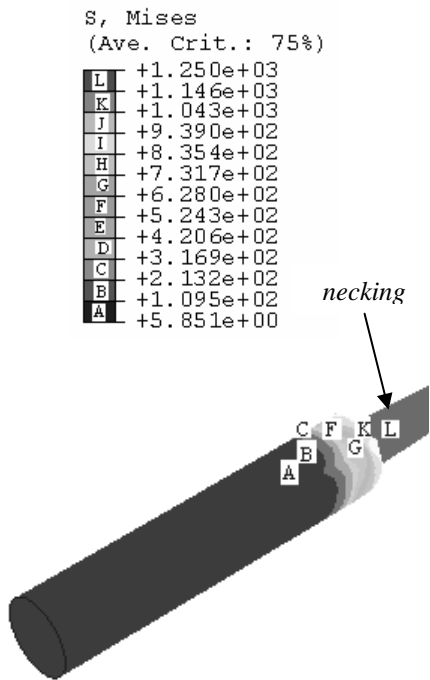


Figure 4. Simulation results showing the stress contours in tensile test bar at strain = 3, strain rate = 10 s^{-1} and temperature = $400 \text{ }^\circ\text{C}$.

Figure 5 and 6 show stress-strain curves produced by tensile simulations at various temperatures and strain rates. The curves follow typical trend of tensile stress-strain with thermal softening i.e. the stress decreases at increasing temperature. However, the stress values are much higher compared to those in the literature. As shown in Figure 4, the maximum stress induced in the simulated bar at the strain rate 10 s^{-1} and temperature 400°C was 1250MPa while that in the actual test bar at strain rate 0.0003 s^{-1} and temperature 400°C was found to be 500MPa [16]. This could be because the applied strain rates in tensile simulation were very much higher than those in practical tensile test. In this comparison, simulated strain rate is 10^4 times higher than the practical strain rate. It may also be the reason that the material is sensitive to strain rate at low strain rates.

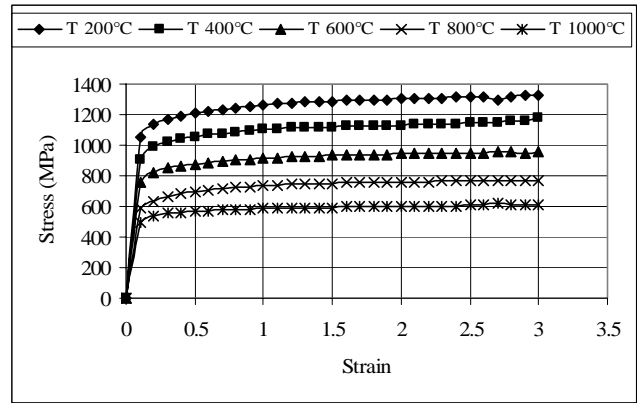


Figure 5. Stress-Strain curves of Ti6Al4V obtained by tensile test simulation with strain rate 10 s^{-1} .

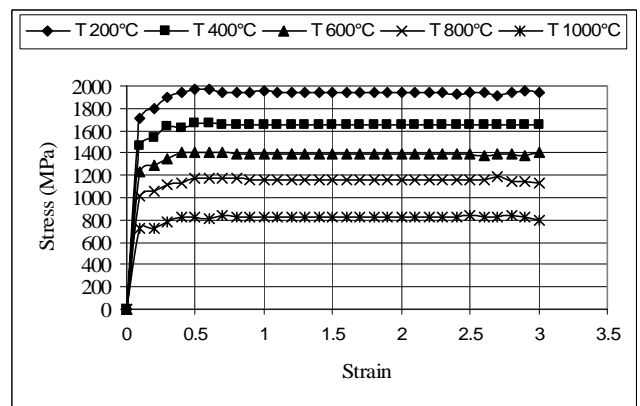
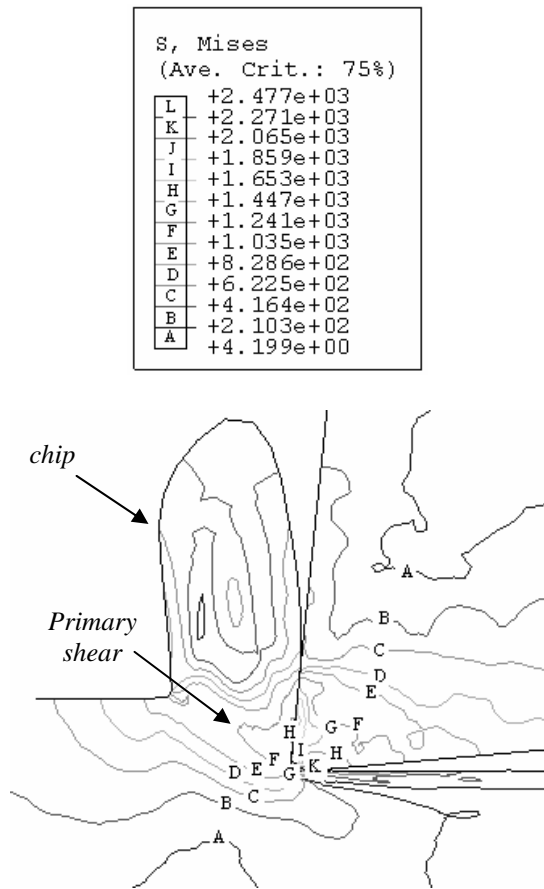


Figure 6. Stress-Strain curves of Ti6Al4V obtained by tensile test simulation with strain rate $100,000 \text{ s}^{-1}$.

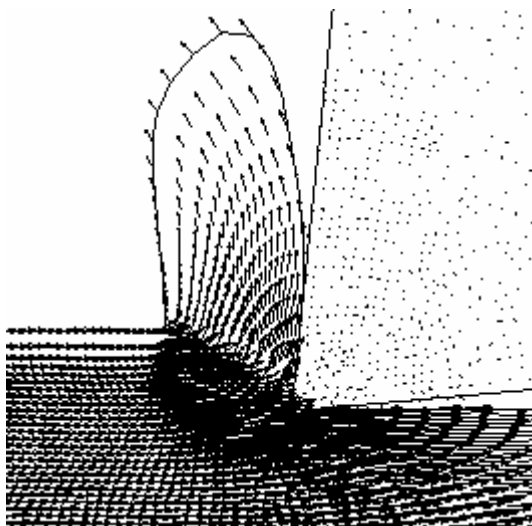
Subsequently the stresses in tensile simulation were compared to those in the cutting simulation. Figure 7(a) illustrates chip flow on the tool rake face during cutting simulation. The smooth flow of the chip on the tool surface was affirmed by Figure 7(b) showing that the material flow lines were consistent with each other as the chip was formed. The temperature distribution induced in the simulated chip was shown in Figure 8. To compare the stresses in tensile and cutting simulations, one element in the primary shear zone which had temperature of about 400°C was picked up and stress-strain was plotted as shown in figure 9. The stress-strain curve shows similar trend with that of tensile test simulation. The maximum stress induced in the primary shear was 1300MPa and the difference in stress between the tensile and cutting simulation was $< 5\%$.

The cutting model was validated by comparing the chip-tool contact length in simulation and experiment. Figure 10 describes the SEM picture of the PCD tool rake face after machining titanium alloy. The white layer represents work material adhered to the tool as the chip flows over it. The width of the white layer represents an approximate chip-tool contact length which is found to be

0.36 mm. The simulation result provides 0.35 mm; so the difference is <3%.



(a)



(b)

Figure 7. (a) Stress contours and (b) material flow lines in cutting simulation.

TEMP (Ave. Crit.: 75%)	
L	+7.001e+02
K	+6.668e+02
J	+6.334e+02
I	+6.001e+02
H	+5.667e+02
G	+5.334e+02
F	+5.001e+02
E	+4.667e+02
D	+4.334e+02
C	+4.000e+02
B	+3.667e+02
A	+3.333e+02

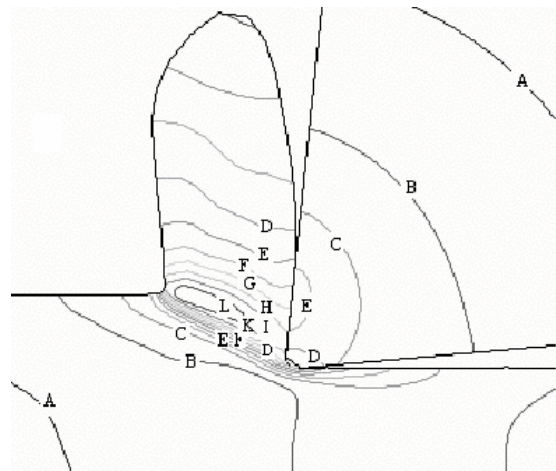


Figure 8. Temperature distribution in cutting simulation.

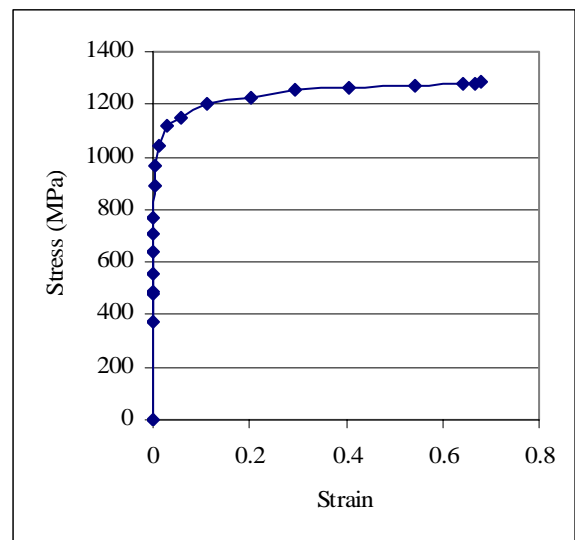


Figure 9. Stress-strain curve of Ti6Al4V obtained from cutting simulation.

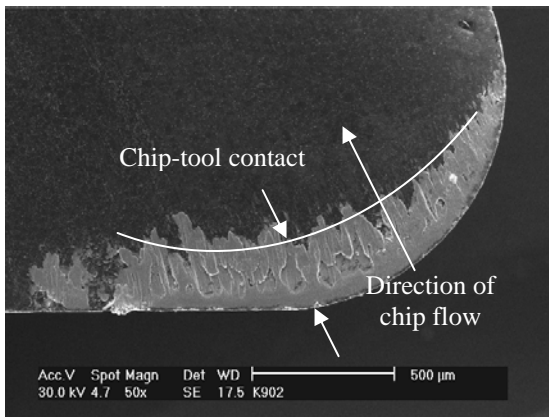


Figure 10. The rake face of the PCD tool after machining Ti6Al4V to verify the chip-tool contact length shown in simulation [Note: The white layer is Ti6Al4V adhered to the rake face near cutting edge].

6. Conclusion

1. The flow stress model with the selected material parameters for Ti6Al4V obtained from the literature is capable of showing material flow at very high strains, strain rates and temperatures.
2. Material flow of Ti6Al4V is more sensitive to temperature than to the strain rate.
3. The proposed tensile simulation technique can be applied to verify the reliability of a material flow stress data.

Acknowledgement

The authors would like to express their sincere thanks to the Ministry of Science and Technology and Innovation, Malaysia for funding this research with IRPA vote number 74218. Special thanks also to Liew and Pang for their consistent help and courage in developing finite element model.

References

- [1] Usui, E., and Shirakashi, T. 1982. Mechanics of Machining from Descriptive to Predictive Theory, on the art of cutting metals- 75 years later. *ASME-PED* 7: 13-15.
- [2] Iwata, K.; Osakada, A.; and Terasaka, Y. 1984. Process Modeling of Orthogonal Cutting by Rigid-plastic Finite Element Method. *Journal of Engineering Material Technology* 106: 132-138.
- [3] Strenkoski, J. S.; and Carroll, J. T. 1985. A Finite Element Model of Orthogonal Metal Cutting. *Journal of Engineering for Industry* 107: 347-354.
- [4] Marusich, T. D.; and Ortiz, M. 1995. Modeling and Simulation of High-speed Machining. *International Journal for Numerical Methods in Engineering* 38 (21): 3675-3694.
- [5] Ceretti, J. E.; Fallbehmer, P.; Wu, W. T.; and Altan, T. 1996. Application of 2D FEM on Chip Formation in Orthogonal Cutting. *Journal of material processing Technology* 59: 169-181.
- [6] Movahhedy, M.; Gadala, M. S.; and Altintas, Y. 2000. Simulation of Orthogonal Metal Cutting Process using an Arbitrary Lagrangian-Eulerian Finite Element Method. *Journal of material processing Technology* 103: 267-275.
- [7] Movahhedy, M., Altintas, Y., and Gadala, M. S. 2002. Numerical Analysis of Metal Cutting with Chamfered and Blunt tools. *ASME* 124 (222): 178-188.
- [8] Pantale, O., Bacaria, J.-L. Dalverny, O., Rakotomalala, R., and Caperaa, S. 2004. 2D and 3D Numerical Models of Metal Cutting with Damage Effects. *Computer methods in applied mechanics and engineering* 193: 4383-4399.
- [9] Wu, H. Y.; Lee, W. B.; Cheung, C. F.; To, S.; and Chen, Y. P. 2005. Computer Simulation of Single-point Diamond Turning using Finite Element Method. *Journal of material processing Technology* 167: 549-554.
- [10] Childs, T. H. C. 1998. Material Properties Needs in Modeling Metal Machining. *Machining Science and Technology* 2 (2): 303-316.
- [11] Hamann, J. C., Meslin, F., and Sarkulvanich, J. 2002. Criteria for the Quality Assessment of Constitutive Equations Dedicated to Cutting Models. *Machining Science and Technology* 6 (3): 331-351.
- [12] Partchapol, S., and Taylan, A. 2005. Effects of Flow Stress and Friction Models in Finite Element Simulation of Orthogonal Cutting – A Sensitivity Analysis. *Machining Science and Technology* 9: 1-26.
- [13] Ozel, T.; and Zeren, E. 2004. Determination of Work Material Flow Stress and Friction for FEA of Machining using Orthogonal Cutting Test. *Journal of Material processing Technology* 153-154: 1019-1025.
- [14] TiMET AUTOMOTIVE USA. 2002. Technical Data.
- [15] ABAQUS. 2004. User's Manual Version 6.5, ABAQUS Inc., USA.
- [16] Marmy, P.; Leguey, T.; Belianov, I.; and Victoria, M. 2000. Tensile and Fatigue Properties of Two Titanium Alloys as Candidate Materials for Fusion Reactors. *Journal of Nuclear Materials* 283-287: 602-606.

Microstructure and Creep Behaviour of AS-Cast Binary Two Phase Gamma TiAl

¹E. Hamzah, ¹M. Kanniah and ²M. Harun

¹ Faculty of Mechanical Engineering, Universiti Teknologi Malaysia,
81310 Skudai, Johor Malaysia

Tel: 07-5534563

Fax: 07-5566159

e-mail: esah@fkm.utm.my

² Industrial Technology Division, Malaysian Institute for Nuclear Technology Research,
Bangi, 43000 Kajang, Selangor Malaysia

ABSTRACT

The microstructure and creep behaviour of as-cast binary two phase gamma TiAl with nominal composition of Ti-48at.%Al was investigated. Tensile creep experiments were performed in air at temperatures from 600-800°C and initial stresses ranging from 150 to 180MPa. Stress exponent and activation energy were both measured. Data indicates that the alloy exhibits steady state creep behaviour. Microstructure before and after creep deformation were examined using scanning electron microscopy techniques.

KEYWORDS: microstructure, creep, TiAl

1.0 INTRODUCTION

The aircraft industry has realized a considerable research effort into the development of structural materials that combine low density with high-temperature applications. This research has led to titanium-aluminum alloys based on the gamma, TiAl phase because of their high strength to weight ratio, high Young's moduli and good oxidation resistance. These alloys also exhibit high strength and creep resistance at temperatures up to 800°C associated with a high activation energy for diffusion. This high energy barrier has its origin in the strong Ti-Al bond. However, this rigid bond structure leads to a limited number of available slip systems. As a result, this alloy exhibits brittle fracture and poor ductility at room temperature (2 – 3%) [1].

Their low density combined with a balance of properties (high temperature strength, toughness and oxidation resistance) makes them viable for certain structural applications (gas turbine components and automotive exhaust valves) up to 700°C or more. Extending the use of these alloys to a wider range of high-temperature structural applications requires improvements in oxidation and creep resistance [2]. All of the envisaged applications for near gamma TiAl intermetallics are within the temperature-stress regime in which creep deformation may occur and creep resistance represents a major obstacle to increasing the service temperature for near gamma TiAl [3].

Although extensive research had been done regarding the creep properties of Ti-48at.%Al, the data obtained is not really appropriate or relevant for real application analysis or extrapolation. Firstly, most of the data obtained for creep studies are from compression

testing [4,5,6] due to its simpler sample preparation. Although the results are likely to be similar under tension, for practical applications, the tensile loading is more relevant and such data will be required for better evaluation of the material capability.

Secondly, creep tests are usually conducted at levels of stress that are higher than the likely real application design stress levels [4,7]. There is a strong need to measure creep response at low stresses and over long times. Furthermore, at higher stress levels, TiAl usually exhibit minimum creep behaviour in contrast with the steady state behaviour exhibited during application.

The stress exponent and activation energy obtained in most of the studies are by assuming a power-law dependence of the creep strain rates on stress and temperature of the form $\dot{\epsilon}_{ss} = A_0 \sigma^n \exp(-Q/RT)$, where A_0 is a material constant, σ the applied stress, Q the activation energy of the creep process, R the perfect gas constant and T the absolute temperature. This equation is, however, only valid in the case of steady-state creep, where the rate of dislocation generation, i.e. hardening, is assumed to be balanced by the rate of dislocation annihilation, i.e. recovery. Most of the creep study at high stresses had analyzed minimum creep behaviour with power law equations which is not actually suitable [8]. Therefore, those data's are not reliable.

Third, detail TEM analysis of crept samples is still lacking. A detail TEM studies is needed to understand the poorly understood creep mechanism of two phase gamma titanium aluminide rather than fitting the mechanical data to the existing creep mechanisms.

The aim of the present work is to study the creep behaviour of as-cast Ti-48at.%Al at real service temperature-stress level with tensile creep testing. The

data obtained will reflect the real behaviour of the alloy at service environment.

2. 0 MATERIALS AND METHOD

The materials investigated in this study have the nominal composition of Ti-48at.%Al. The alloys were produced by plasma melting casting in the form of 2kg buttons in IRC Birmingham. The alloys will be referred in atomic percentage hereafter.

Initial as-cast microstructure was examined using optical and Phillips XL40 scanning electron microscope in back scattered electron (BSE) mode.

Flat samples with dimension, 3mm x 2mm with 15mm gauge length were prepared for the creep test. Prior to creep test, the test pieces were ground along the gauge length in the longitudinal direction to prevent premature crack initiation at surface defects during creep.

The constant load tensile creep test was conducted using Mayes TC 20 High Temperature Creep Testing Machine, consisting three zone temperature furnace and a loading lever arm ratio of 10:1 in air. The test temperatures were maintained constant with a precision of $\pm 1^{\circ}\text{C}$ and monitored for 2 hours before the test. The temperature was monitored during the creep test by using a thermocouple in contact with the gauge section of the sample. The sample was allowed to soak at test temperature for 1 hour before the load was applied. The strain was measured using a HTD type extensometer which is attached with two linear voltage displacement transducers (LVDT) with resolution of $\pm 1\mu\text{m}$ and a data logger. The strain was measured for every 1 second intervals at the start of the creep test, increasing to a maximum interval of 15 minutes. The test was interrupted after 20 hours and the samples were left to cool in air to room temperature.

3. 0 RESULTS AND DISCUSSION

3.1 Initial As-Cast Microstructure

The as-cast microstructure prior to creep testing was examined carefully and quantitatively in order objectively to distinguish later any creep induced microstructural changes. Figure 1(a) and (b) shows the optical and SEM-BSE micrograph of typical nearly lamellae microstructure in the gauge region of creep samples of as-cast Ti-48Al. The microstructure consists of coarse lamellae grains with an average grain size of 500 to 800 μm and 2 to 5% of fine gamma grains at the grain boundaries. The lamellae structure consists of alternating lathes of the α_2 -Ti₃Al and γ -TiAl phases. The α_2 laths are bright in contrast due to their higher titanium content. The lamellae structure results from the solid state phase transformation of the primary disordered α dendrites. The

single-phase γ regions surrounding the lamellae grains result from the transformation of the aluminum rich interdendritic melt, which recrystallize to give rise to the fine γ grains [8]. The serrated grain boundary is caused by incursions of lamellae into the neighbouring grain [9].

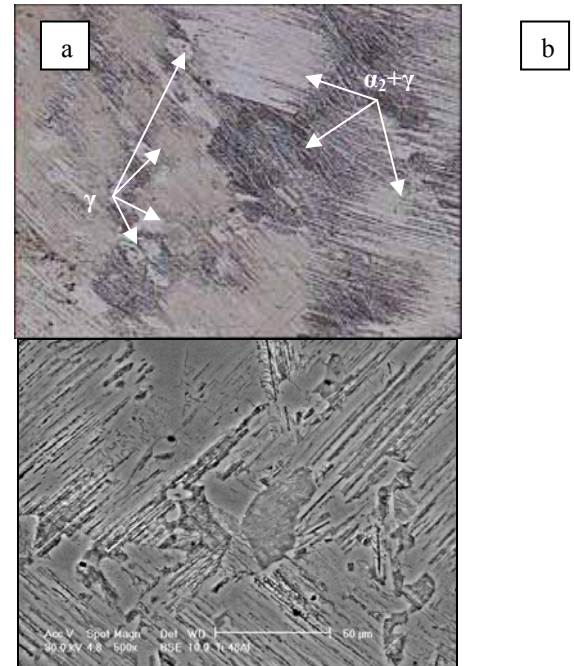


Figure 1 Initial as-cast microstructure of Ti-48Al (a) Optical (50x) and (b) SEM-BSE (500x)

3.2 Creep Behaviour

All the creep curves obtained in this study have the same basic shape, as shown in figure 2 and 3. Figure 2(a) and (b) illustrates the creep strain (%) versus time (hrs) and strain rate (hr^{-1}) versus time (hrs) for creep tests at temperatures ranging between 600-800 $^{\circ}\text{C}$ and identical initial stress of 180MPa respectively. Figure 3(a) and (b) illustrates the creep strain (%) versus time (hrs) and strain rate (hr^{-1}) versus time (hrs) for creep tests at identical temperature 800 $^{\circ}\text{C}$ and initial stresses range from 150-180MPa respectively. The creep strain increases rapidly in the primary creep regime until strain rate gradually falls to a steady-state value. During the primary creep stage the creep rate decreases with increasing strain. In contrast to the previous results [10,11], where 'minimum creep regime' is reported, no minimum creep behaviour is observed over the studied temperatures and stresses investigated.

Analysis on the creep curves of figure 2(a) shows Ti-48Al exhibit no instantaneous creep strain at 600 and 700 $^{\circ}\text{C}$ with initial stress of 180MPa. Both these creep curves has almost identical shape. Figure 2(b) shows that Ti-48Al exhibits steady state behaviour at this condition of temperature and stress. However, Ti-48Al exhibits high instantaneous creep strain, 0.6% at 800 $^{\circ}\text{C}$ and initial stress of 180MPa. Creep curves in figure 3(a) shows that creep behaviour of Ti-48Al is very sensitive to stress. This is due

to the total creep strains for 20 hours with initial stress of 180MPa is significantly higher than the total creep strains with initial stress of 150MPa. Ti-48Al exhibits 0.53% and 0.6% instantaneous creep strain when loaded with initial stress of 150 and 180MPa at 800°C respectively. Ti-48Al showed steady state behaviour at both these conditions also as shown in figure 3(b).

The dependence of the steady state creep rate on temperature and stress can be formulated by the power law creep equation (1),

$$\dot{\epsilon}_{ss} = A \sigma^n \exp(-Q/RT) \quad (1)$$

where $\dot{\epsilon}_{ss}$ is steady state creep rate, A is material constant, σ is applied stress, n is stress exponent and Q is the activation energy. Figure 4(a) shows the variation of the steady state creep rate with the applied stress. Using linear regression analysis, the stress exponent n is determined to be ~ 3 over the studied stress range. Figure 4(b) shows the temperature dependence of the steady state creep rate in the form of Arrhenius diagram. From this figure, the activation energy for creep Q calculated for three different temperatures with initial stress of 180MPa is 20.4kJ/mol. However, fitting these mechanical data to the existing creep mechanisms without detail understanding on the behaviour of the material as a function of microstructural variation or creep regime would not be appropriate. A detailed TEM study is necessary for understanding the real creep mechanisms.

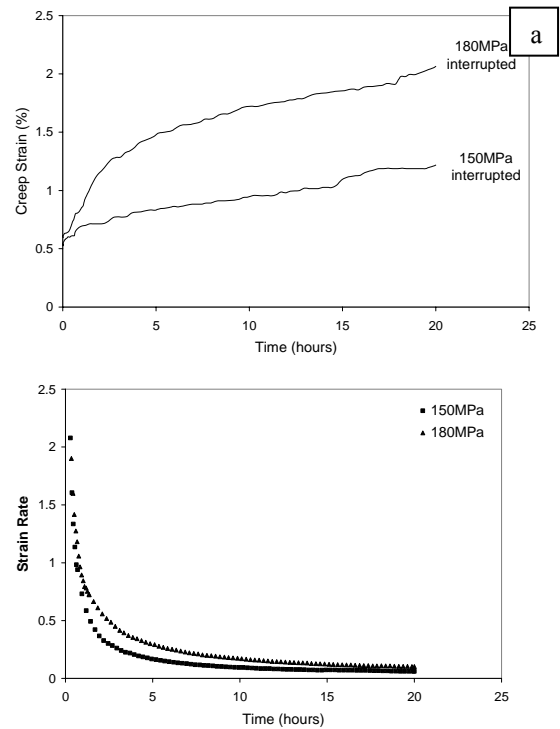


Figure 3 Creep curves of Ti-48Al at identical temperature 800°C and initial stresses ranging between 150-180MPa (a) Creep strain (%) vs Time (hours) and (b) Strain rate (hr⁻¹) vs Time (hours)

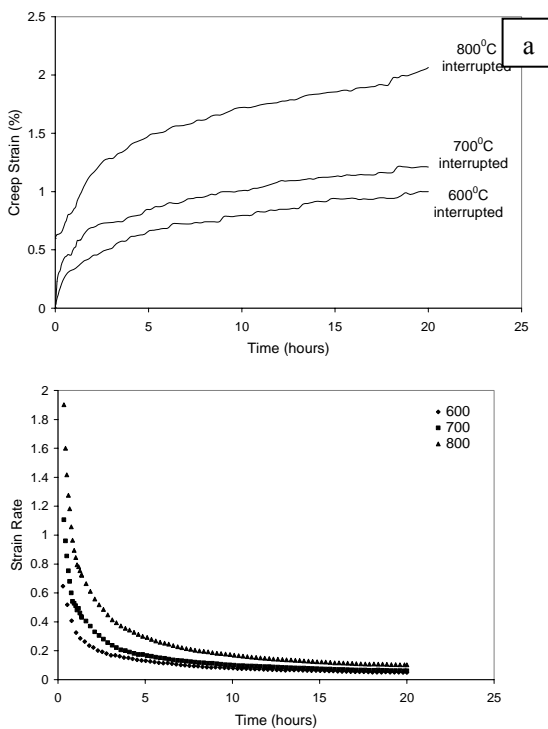


Figure 2 Creep curves of Ti-48Al at temperatures ranging between 600-800°C and identical initial stress of 180MPa (a) Creep strain (%) vs Time (hours) and (b) Strain rate (hr⁻¹) vs Time (hours)

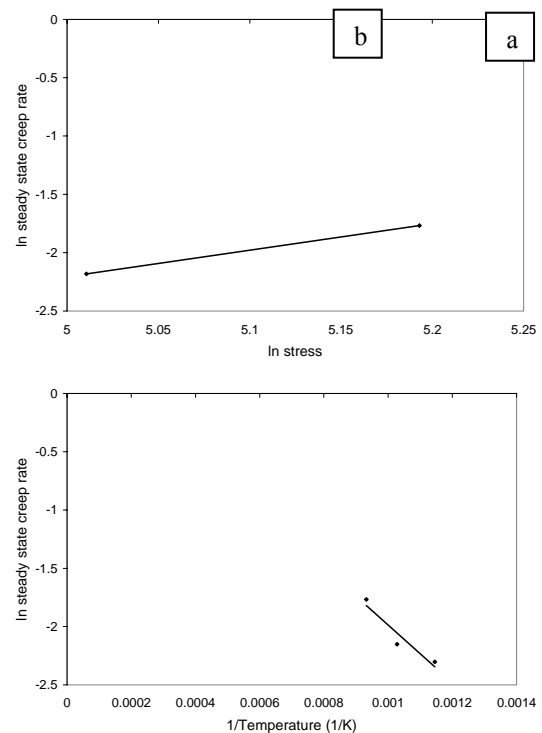


Figure 4 Dependence of steady state creep rate on (a) Stress (ln $\dot{\epsilon}_{ss}$ vs ln σ) and (b) Temperature (ln $\dot{\epsilon}_{ss}$ vs 1/T)

3.3 Crept Microstructure

The main two factors which determine the shape of creep curve besides development of twins and dislocations are investigated here. The factors are microstructure instabilities and creation and growth of creep cavities. Figures 5(a)-(d) show microstructures of crept samples of Ti-48Al which is interrupted after 20 hours of creep deformation at temperatures of 600-800°C and initial stresses of 150-180MPa. The microstructural changes during steady state creep regime are reflected by dynamic recrystallization and spheroidization of the α_2 laths. Considerable amount of equiaxed and fine γ grains extended to the whole cross section in all the four condition as shown in figure 5(a)-(d). The fine γ grains appear to be due to dynamic recrystallization during creep deformation. The lamellae adjacent to dynamic recrystallized γ grains also experience dynamic recrystallization and the lamellae structure in these regions is no longer preserved as illustrated in figures which shows the very fine scale structure obtained in such areas (refer spot 1, 2, 3 and 4 in the figures). Figure 5(a) also revealed that the lamellae structure is partially recrystallized. In comparing figures 5(a)-(d) with the initial microstructure, it is readily seen that the contrast of the lamellae is no longer uniform. In addition with dynamic recrystallization, Ti-48Al suffers from formation of creep cavities and spheroidization of α_2 lamellae too. Representative micrograph showing formation of creep cavities at grain boundaries and spheroidization of α_2 lamellae in Ti-48Al are shown in figures 6(a) and 6(b) respectively.

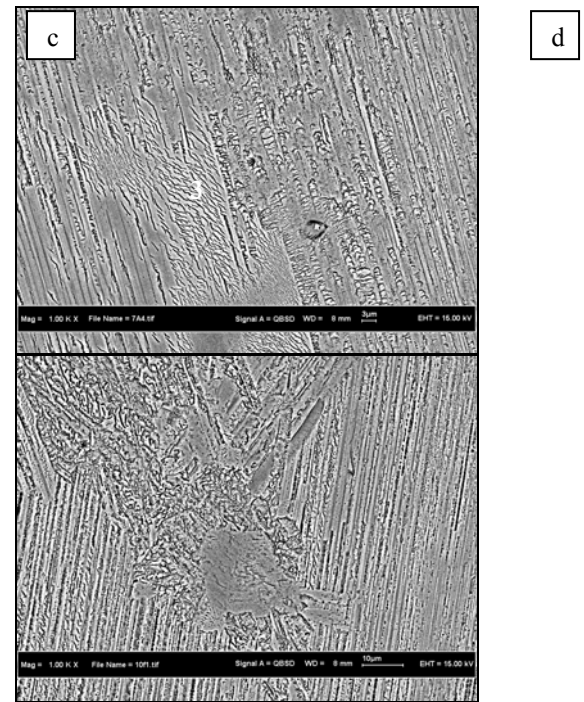
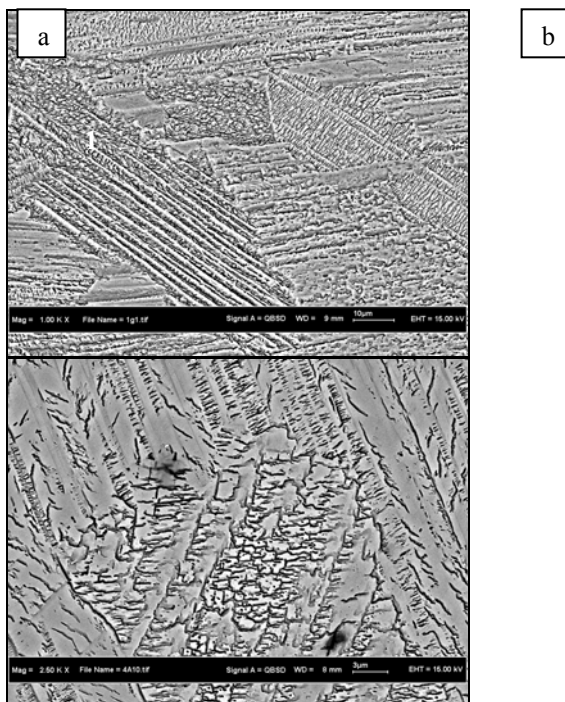


Figure 5 Microstructure of Ti-48Al-2Cr after 20 hours of creep deformation at (a) 800°C and 150MPa (1000x), (b) 800°C and 180MPa (2500x), (c) 700°C and 180MPa (1000x), and (d) 600°C and 180MPa (1000x)

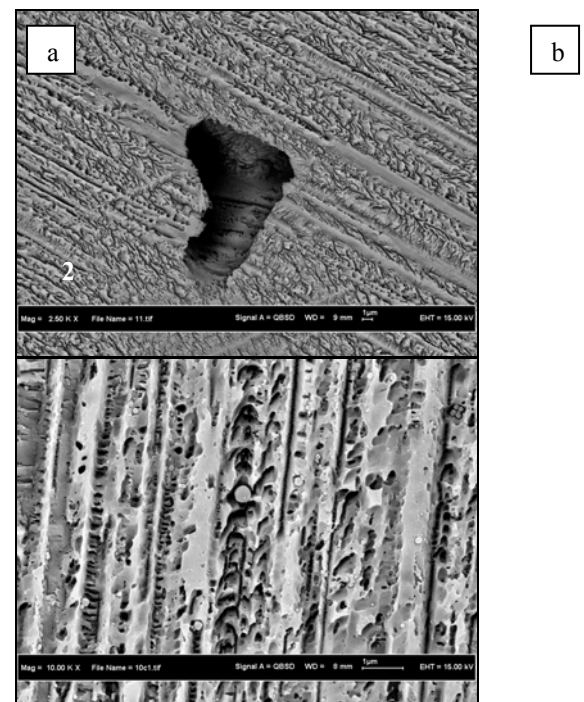


Figure 6 SEM-BSE micrograph showing (a) Creep cavities after 20 hours of creep deformation at 800°C and initial stress of 150MPa (2500x) and (b) Spheroidization α_2 lamellae after 20 hours of creep deformation at 600°C and initial stress of 180MPa (10000x)

4.0 CONCLUSIONS

1. As-cast Ti-48Al exhibits a nearly lamellar microstructure consisting large volume fraction of lamellae regions with grain size from 500-800 μ m and 2-5% of fine γ regions.
2. Ti-48Al exhibited steady state behaviour over the temperatures and stresses investigated.
3. During creep of Ti-48Al alloy, the steady state creep rate is found to depend on the applied load and temperature. The measured power law stress exponent for steady state creep rate is found to be close to 3 and the apparent activation energy for creep is calculated to be 20.4kJ/mol.

ACKNOWLEDGEMENTS

The authors would like to thank the Malaysian Ministry of Science, Technology and Innovation (MOSTI) for the research fund under the IRPA program (Project No. 09-02-06-0002 EA-002).

REFERENCES

1. Jimenez, J.A., Wesemann, J. and Frommeyer, G. 1998. High-Temperature Deformation Behavior of the Intermetallic Ti-47 At. Pct Al-3 At. Pct Cr Alloy. *Metallurgical and Materials Transactions A*. 29A(5): 1425-1433.
2. Parthasarathy, T.A., Subramaniam, P.R., Mendiratta, M.G. and Dimiduk, D.M. 2000. Phenomenological Observations of Lamellar Orientation Effects on the Creep Behavior of Ti-48at.%Al PST Crystals. *Acta Materialia*. 48: 541-551.
3. Chen, W.R., Triantafillou, J., Beddoes, J. and Zhao, L. 1999. Effect of Fully Lamellar Morphology on Creep of a Near γ -TiAl Intermetallics. *Intermetallics*. 7: 171-178.
4. Perdrix, F., Trichet, M.F., Bonnetien, J.L., Cornet, M. and Bigot, J. 2001. Relationships between Interstitial Content, Microstructure and Mechanical Properties in Fully Lamellar Ti-48Al Alloys, with Special Reference to Carbon. *Intermetallics*. 9: 807-815.
5. Lu, M. and Hemker, K.J. 1997. Intermediate Temperature Creep Properties of Gamma TiAl. *Acta Materialia*. 45(9): 3573-3585.
6. Gouma, P.I., Davey, S.J. and Loretto, M.H. 1998. Microstructure and Mechanical Properties of a TiAl-Based Powder Alloy Containing Carbon. *Materials Science and Engineering Part A*. A241:151-158.
7. Sun, F.S., Cao, C.X., Kim, S.E., Lee, Y.T. and Yan, M.G. 2001. Alloying Mechanism of Beta Stabilizers in a TiAl Alloy. *Metallurgical and Materials Transactions A*. 32A: 1573-1589.
8. Es-Souni, M., Bartels, A. and Wagner, R. 1993. Creep Deformation and Creep Microstructures of a Near γ -TiAl Alloy Ti-48Al-2Cr. *Materials Science and Engineering*. A171: 127-141.
9. Herrouin, F., Hu, D., Bowen, P. and Jones, I.P. 1998. Microstructural Changes During Creep of a Fully Lamellar TiAl Alloy. *Acta Materialia*. 46:4963-4972.
10. Kartikeyan, S., Viswanathan, G.B., Gouma, P.I., Vasudevan, V.K., Kim, Y.W. and Mills, M.J. 2002. Mechanisms and Effect of Microstructure on Creep of TiAl-Based Alloys. *Materials Science and Engineering Part A*. A329-331: 621-630.
11. Dlouhy, A., Kucharova, K. and Brezina, J. 2001. Dislocation Slip and Deformation Twinning Interplay During High Temperature Deformation in γ -TiAl Base Intermetallics. *Materials Science and Engineering Part A*. A319-321: 820-826.

Performance of wiper coated carbide tool when turning hardened stainless steel

M. Y. Noordin*, D. Kurniawan, S. Sharif, Y. C. Tang

Faculty of Mechanical Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5534697, Fax: +60-7-5566159, E-mail: noordin@fkm.utm.my

Abstract

Hard turning has been widely applied as finish machining in parts manufacturing. The application of hard turning method is in line with the development of cutting tools. Coating techniques are part of the development that is intended to fortify cutting tools to more extreme conditions. New coating over improved tool substrate potentially allows it to be employed for hard turning. Various ways have been developed to improve productivity of cutting tools. Wiper (multi radii) geometry has been provided by tool manufacturers for that purpose. Tool with more wear resistant and carbide grades and enhanced geometry offer the option to double current feed rate and still achieve surface roughness value comparable to that produced using conventional inserts. This study investigated the performance of wiper coated carbide tool when hard turning stainless steel. Wiper coated carbide tool successfully performed hard turning of stainless steel of 47-48 HRC. By varying the cutting speed and feed rate, the tool life increased with decreasing speed and feed rate. Further visual observation using optical and scanning electron microscope on the worn tool revealed that wear occurred at the flank and rake face of the tool. Abrasion was the main wear mechanism involved.

Keywords: Hard turning; Wiper insert; Coated carbide; Tool life; Wear mechanism

1. Introduction

Machining is major metal shaping process in mechanical manufacturing industry. Hardened parts are commonly manufactured by involving heat treatment process to control the hardness during and after processes and finish machining process. Hard turning of machine parts holds considerable potentials since it is an effective means of increasing productivity in manufacturing hardened parts. Hard turning has been explored as an alternative to grinding for finish machining, by cutting parts in their hardened state instead of machining a product in the soft state, hardening the part by heat treatment, and then providing the required dimensional accuracy by grinding [1].

The types of materials being cut by hard turning method are growing in numbers and applications. Steels are the common materials subject to hard turning. Stainless steel is among the materials being investigated to employ hard turning method due to its wide application in automotive, construction, food, chemical, oil and gas, automotive, and tool and die industries.

The current study used martensitic stainless steel as the material being hard turned to its final dimensions. Martensitic stainless steel is a type of stainless steel that has relatively high carbon content (0.1 to 1.2%), containing between 12 and 18% chromium. This stainless steel is hardenable by heat treatment and therefore high strength and hardness levels can be achieved [2].

The applications of hard turning are made possible by the development of tool materials and coating technology.

TiAlN-based coating is a type of coating that shows high hardness, good thermal stability, and oxidation resistance which enhance tool performance in many metal cutting applications, and suitable for high speed and dry machining. Coated cemented carbide tools constitute a combination of suitable properties (i.e., adequate fracture toughness, stable thermal qualities, etc.) which can ensure reasonable tool life at minimal cost per cutting edge relative to ceramic tools [3-6]. However, the use of coated carbide as a tool material has not been explored extensively in hard turning.

For the purpose of providing productivity acceleration in implementing inexpensive hard turning method, this current study utilized carbide inserts with special multi radii (wiper) geometry. Since the cost of machining is very strongly dependent on the rate of metal removal, and costs may be reduced by increasing the cutting speed and/or the feed rate, this geometrically modified insert has the potential to increase quality and productivity. These improvements of more wear resistant carbide grades and enhanced tool geometries can double current feed rate and still achieve surface finishes comparable to conventional inserts, as shown in Fig. 1. The principle behind wiper geometry is that inserts can be made to fulfill two manufacturing requirements at once, high productivity and surface finish. If surface finish is the most important consideration, then the same feed can be maintained. If surface finish is not critical, then the same finish can be achieved with the current tool operating at double the previous feed rate [7,8].

For more than 20 years, wiper technology has been applied to improve surface finish for milling operations, but development of the wiper insert for turning applications is relatively new. It may be interesting to note that wiper inserts in turning are not yet utilized in as many finish turning

* Corresponding Author. E-mail: noordin@fkm.utm.my,
Tel: +60-7-5534697, Fax: +60-7-5566159

applications as they could possibly be [7].

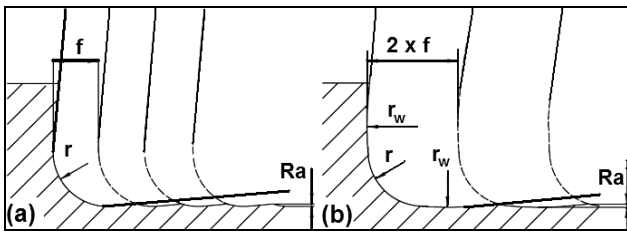


Fig. 1. Comparison between (a) conventional inserts and (b) wiper, where f = feed, r = corner radius, r_w = wiper radius, and Ra = surface finish.

2. Experimental Setup

The experiments were performed on a 2-axis CNC lathe machine set for dry cutting (without the use of coolant). The selected conditions/values of each of these variables are given below.

(a) *Workpiece material*: STAVAX, an AISI 420 (modified) stainless steel bar of hardness 47-48 HRC with 90 mm diameter and 250 mm length. The chemical composition of this workpiece is 0.38% C, 0.9% Si, 0.5% Mn, 13.6% Cr, and 0.3% V.

(b) *Tool insert and holder*: Kennametal's KC5010FW coated carbide (PVD TiAlN coating over a deformation-resistant unalloyed carbide substrate, with coating thickness of 3.0 to 3.5 μm) as the tool insert. The used insert has an ISO designation of CNMG 120408. The tool holder used was ISO designated as MCLNL 1616-H12.

(c) *Tool geometrical parameters*: Multi radii (wiper) geometry inserts, modified form of 0.8 mm nose radius conventional insert, with -5° rake angle, 5° relief angle, and 5° side cutting edge angle.

(d) *Cutting conditions*: Cutting speeds of 100, 130, and 170 m/min were used with feed rates of 0.125, 0.16 and 0.25 mm/rev by keeping a constant depth of cut of 0.4 mm.

For each test condition, the average width of flank wear land (VB_C) and maximum width of flank wear land (VB_{Max}) were measured after machining for a predetermined time interval.

3. Tool Life and Wear Mechanisms

3.1. Tool Life

The tool life of the wiper coated carbide inserts was determined by the criteria of $VB_{max} = 0.14$ mm or when severe tool chipping (catastrophic failure) occurred and the surface finish of the machined surface of not higher than $Ra = 1.6$ μm .

Tool life variation with cutting speed and feed rate can be seen in Fig. 2. As expected, the tool life of wiper coated carbide insert decreased with increasing cutting speed and feed rate. The highest tool life was obtained at a speed of 100 m/min and feed rate of 0.125 mm/rev, i.e. combination of low speed and low feed rate. A trend can also be seen that increasing the feed rate at the same cutting speed resulted better tool life than increasing the cutting speed by keeping the feed rate constant (e.g. tool life of 100 m/min speed and 0.16 mm/rev feed was better than that of 130 m/min speed

and 0.125 mm/rev feed). This may lead to the suggestion that for this range of cutting conditions, increasing feed rate is a preferred option than to increase the cutting speed when time of delivery constraints the machining process time.

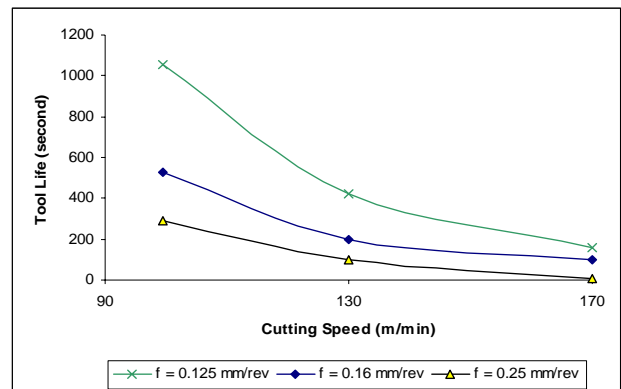


Fig. 2. Tool life of wiper coated carbide inserts.

By considering the tool life values, the use of the combination of 170 m/min and 0.25 mm/rev cutting speed and feed rate, i.e. high speed-high feed, was not feasible. Using such a combination would be a burden to implement at manufacturing workshops since tool changing would be extremely frequent and low productivity of the insert used resulted in an increased cost per part.

3.2. Wear Mechanisms

The wear that was clearly observed on the tool was flank wear and crater wear at the nose region. By considering tool wear progress in Fig. 3, it can be noticed that there were three stages occurred: initial stage, gradual stage, and abrupt stage. In Fig. 3, cutting conditions at low speeds showed longer gradual stage, unless at high feed where medium speed showed more obvious gradual stage. These typical behaviors were in agreement with those observed previously [9].

Previous observations reported that flank wear occurred on the tool flank and was generally attributed to rubbing of the tool with workpiece at the interface, causing abrasive and/or adhesive wear. Flank wear in the tool can be characterized by grooves and ridges in the direction of tool sliding against a newly machined surface of the workpiece or chip sliding against the tool rake. The severity of abrasion may increase in cases where the workpiece materials contain hard inclusions, such as hard carbide particles or clusters, or when there is debris from the workpiece or the tool at the interface, or by loose hard grains of the tool (which were swept away due to chemical/dissolution/diffusion wear of the binder) [3,6,10,11].

In this study on hard turning of martensitic stainless steel, as can be seen in Fig. 4, grooves existed and even appeared during the initial stage of cutting time and were there until the final stage of cutting. And since martensitic stainless steel consists of hard carbide particles, the grooves appearance might be attributed to the abrasive action of these carbide particles. The appearance of carbide particles can be seen from the microstructure image of the workpiece material in Fig. 5(a).

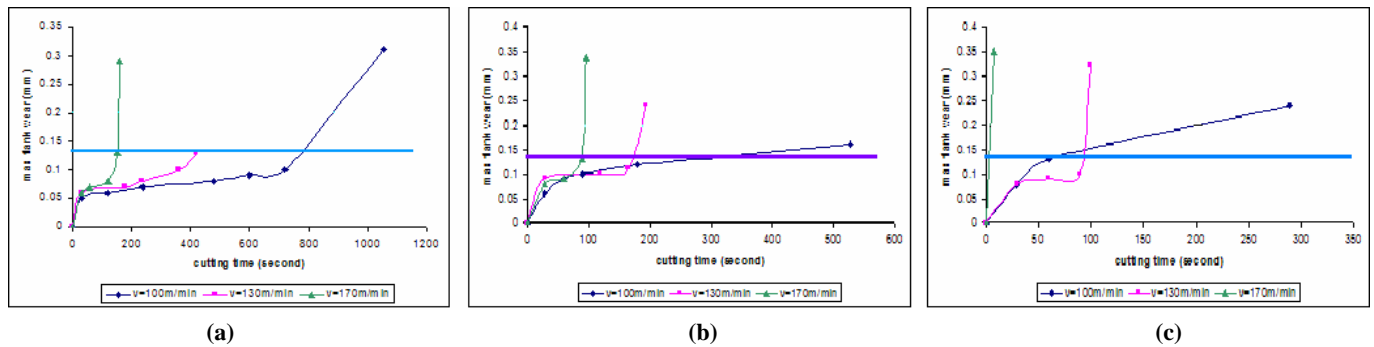


Fig. 3. Flank wear progress as the function of cutting time at 0.4 mm depth of cut and feed rates of (a) 0.125 mm/rev, (b) 0.16 mm/rev, and (c) 0.25mm/rev.

Wear also appeared at the tool's rake face. The rake face of the tool was subjected to abrasion that caused removal of the tool's coating and substrate at region slightly further from the tool edge. From Fig. 6(a) and 7, it is seen that the abrasive action formed a crater and caused the exposure of the tool's substrate. Since the rake face of the tool was where the tool-chip contact occurred, the abrasive action might be caused by hard carbide particles of the chip. Fig. 5(b) shows the appearance of the cross-sectional chip's microstructure that is similar to the workpiece's. Previous studies reported that the most significant factors influencing crater wear were temperature at the tool-chip interface and the chemical affinity between tool and workpiece materials. Also mentioned that this type of wear was more pronounced at high cutting speeds or when there was high temperature at the tool-chip interface, and was accelerated by high chemical affinity between the workpiece and the tool [10,12]. Another factor that should be taken into consideration in determining wear at the tool's rake face was the high force involved in chip generating action, as evidenced by the plastically deformed chip's underside (the bottom of the chip), as in Fig. 5(b). The chip's affected area was approximately until the depth of 13 μm from the base of the underside. The chip's microstructure, appearing two bands on the altered microstructure, also indicated the high temperature generated simultaneously during turning the hardened stainless steel workpiece. As can be seen, the tool's underside until around 10 μm deep consists of thinner grains compared to those of the bulk, that are identical to the workpiece's, with 3 μm dark band intermediating between the two regions.

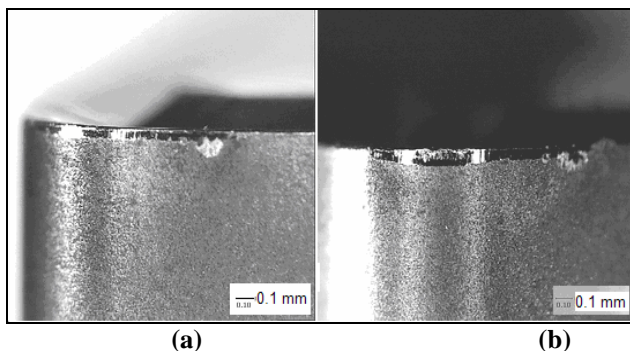


Fig. 4. Images of flank wear on the wiper tool: (a) minor cutting edge view at early stage of cutting time and (b) nose region view prior to the end of tool life.

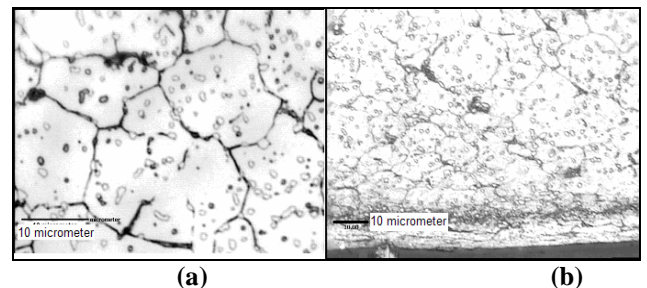


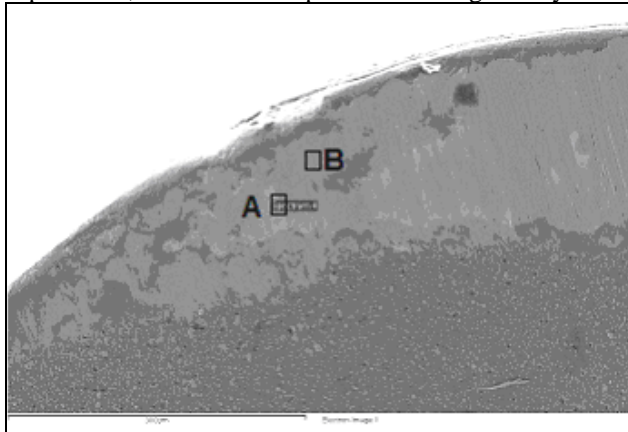
Fig. 5. Microstructure of martensitic stainless steel (a) workpiece (1000X) and (b) chip cross-section (500X) (etched by Vilella's reagent).

From the examination using energy dispersive spectrometry, as shown in Fig. 6, it can be observed that different situations identified at two regions in the tool rake face. Region A showed abrasion of the tool coating since the elements of the tool substrate were exposed, while region B showed workpiece elements adhered to the rake face. The appearance of workpiece elements on rake face might be caused by the resolidified workpiece material after being melted during tool-chip contact. The abrasive action might be caused by hard carbide particles contained in the chip, as can be observed from Fig. 5(b), which were not melted since the carbide's melting temperature is above the cutting temperature generated.

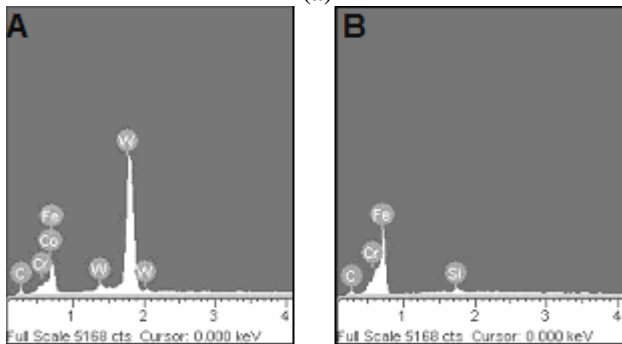
Mentioned by previous reports, the condition of exposed tool substrate due to coating abrasion caused by hard carbide particles in the workpiece material might introduce the formation of crack initiation. Once a crack was initiated, "hammering" action of the fast flowing chip on this weak site might propagate extensive coating abrasion, and followed by severe tool breakage [6,13]. For this hard turning using wiper coated carbide tools, such condition occurred. Most of the tool tips used (seven of nine cutting experiments done, indicated by maximum flank wear value of over 0.14 mm in Fig. 3) failed by severe edge chipping.

Microcracks were clearly identified by examining the rake face after releasing the adhered workpiece material on tool insert just before it failed (Fig. 7). It can be seen that crack occurred both at the coating and at the substrate. The crack of the substrate indicates the damage of the tool tip, which might be the main cause of edge chipping or even catastrophic failure. The dark objects laid on the crater were identified as the coating material. Supporting previous

explanation, the abrasion impacted unhomogeneously over



(a)



(b)

Fig. 6. (a) SEM image of the rake surface of wiper coated carbide inserts at the end of tool life and (b) its related EDS analysis of region A and region B at the rake.

the tool's rake. The unlevelled height of the substrate potentially became the weak sites where cracks initiated or propagated. The cracks at some critical points triggered the tool failure since the tool could no longer stand the high cutting forces generated during hard turning process.

The tool surface might also be swept away by a mechanism termed as workpiece adherence [14]. In Fig. 10, the adhered workpiece material appeared unidirectionally in lighter colour than the coating at the crater. Due to strong adhesion, the workpiece material adhered on the tool surface, accumulated, and, until the size reached critical value, detached and followed by the underlying coating as well as small pieces of WC substrate.

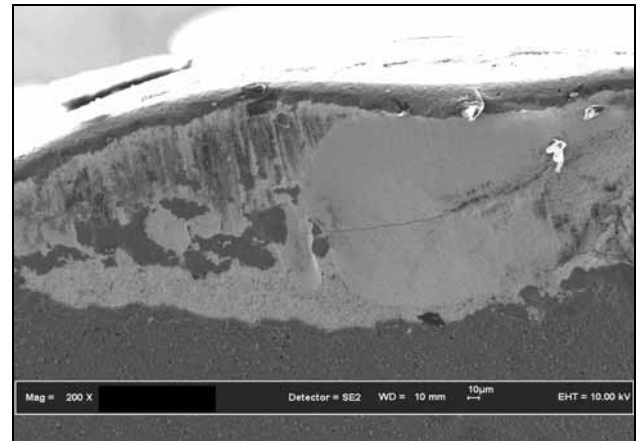


Fig. 7. The rake face of worn wiper coated carbide tool.

3. Conclusion

Hard turning of martensitic stainless steel was done using wiper coated carbide insert at various cutting speeds and feed rates. The cutting condition of low speed and low feed rate performed the longest tool life. Only the combination of high cutting speed and high feed rate that was not applicable for hard turning application. Wear mechanism was mainly abrasion. The abrasive action and plastic deformation, discrete plastic deformation (of the coating), and detached adhered workpiece swept away the tool coating and substrate. Microcracks were also initiated and propagated resulting in severe edge chipping or even catastrophic failure for most of the tools used.

Acknowledgments

Financial support from the Ministry of Science, Technology, and Innovation Malaysia through the IRPA funding vote no. 74268 is acknowledged with gratitude.

References

- [1] Shaw, M. C., and Vyas, A. 1998. Hard turning of Steel. *Technical Paper of the NAMRI-SME* MR98-240: 1-5.
- [2] Sourmail, T., and Bhadeshia, H. K. D. H. 16 Oct 2005. Citing Internet sources URL http://www.msm.cam.ac.uk/phase-trans/2005/Stainless_steels/stainless.html
- [3] Arsecularatne, J. A., Zhang, L. C., Montross, C., and Mathew, P. 2006. On machining of hardened AISI D2 steel with PCBN tools. *Journal of Material Processing Technology*. 171 (2): 244-252.
- [4] Derflinger, V. H., Schütze, A., and Ante, M. 2006. Mechanical and structural properties of various alloyed TiAlN-based hard coatings, *Surface and Coatings Technology*. 200: 4693-4700.
- [5] PalDey, S., and Deevi, S. C. 2003. Single layer and multilayer wear resistant coatings of (Ti,Al)N: a review. *Materials Science and Engineering A*. 342 (1-2): 58-79.
- [6] Ezugwu, E. O., and Olajire, K. A. 2002. Evaluation of

- machining performance of martensitic stainless steel (JETHETE). *Tribology Letters*. 12 (3): 183-187.
- [7] Castner, M. 2000. Get a finer finish. *Quality*. 39(8): 46.
- [8] Trent, E. M., and Wright, P. K. 2000. *Metal Cutting*. 4th ed. Boston: Butterworth-Heinemann.
- [9] Che Haron, C. H., Ginting, A., and Goh, J. H. 2001. Wear of coated and uncoated carbides in turning tool steel. *Journal of Materials Processing Technology*. 116(1): 49-54.
- [10] Senthil Kumar, A., Raja Durai, A., and Sornakumar, T. 2006. The effect of tool wear on tool life of alumina-based ceramic cutting tools while machining hardened martensitic stainless steel. *Journal of Materials Processing Technology*. 173(2): 151-156.
- [11] Poulachon, G., Bandyopadhyay, B. P., Jawahir, I. S., Pheulpin, S., and Seguin, E. 2003. The influence of the microstructure of hardened tool steel workpiece on the wear of PCBN cutting tools. *International Journal of Machine Tools and Manufacture*. 43(2): 139-144.
- [12] Liew, W. Y. H., Lu, Y. G., Ding, X., Ngoi, B. K. A., and Yuan, S. 2004. Performance of uncoated and coated carbide tools in the ultra-precision machining of stainless steel. *Tribology Letters*. 17(4): 851-857.
- [13] Liew, W. Y. H., Ngoi, B. K. A., and Lu, Y. G. 2003. Wear characteristics of PCBN tools in the ultra-precision machining of stainless steel at low speeds. *Wear*. 254(3-4): 265-277.
- [14] Su, Y. L., Liu, T. H., Su, C. T., Yao, S. H., Kao, W. H., and Cheng, K. W. 2006. Wear of CrC-coated carbide tools in dry machining. *Journal of Materials Processing Technology*. 171: 108-117.

Development Of Activity-Based Costing (ABC) Model In Demanufacture Cost Assessments

S.Mugeneswaran a/l Shanmugam ^{1*}, Awaluddin Mohd Shaharoun ²

¹ Faculty of Engineering & Technical,
Open University Malaysia, Jalan Tun Ismail, 50480 Kuala Lumpur
E-mail : muge @ oum.edu.my

² Business And Advanced Technology Centre (BATC),
University Technology of Malaysia, Jalan Semarak, Kuala Lumpur
E-mail : awal @ fkm.utm.my

Abstract

This study focuses on the implementation of ABC and DfD in a demanufacturing case study where the ABC system is used to calculate the product cost of hand phones repaired using a cycle of Plan-Do-Check-Act incorporating ABC. Various manufacturing costs are traced to activity centers by first stage cost assignment and subsequently to product with second stage cost drivers. The critical activities and cost drivers are identified and improved with engineering tools. Subsequently, the new ABC product cost is recalculated. It also proves that changes of cost structure are reactive with cost driver improvements. Based on the findings, costs are simulated for various products in the same environment. With this a better understanding of the disassembly activities and impact of activities on the cost is achieved. As conclusion, the implementation of ABC cost system in a demanufacturing environment able to determine the cost of product accurately, identify critical activities and cost driver for improvement with significant savings.

Key Words : Activity based costing, DfD

1. Introduction

1.1 Activity Based Costing (ABC)

Activity based costing (ABC) was developed by Harvard Business School Professors Kaplan and Cooper and is a costing system that focuses on activities that contributes to the cost of products. It is based that all products consume activities and activities consume resources therefore consume cost (Kaplan, 1984).

An ABC model is capable identifying the activities that will be present in the demanufacturing process of a product, and afterwards assign reliable cost drivers and associated consumption intensities to the activities (Bert Bras, 1995). ABC provides cost information such as the cost drivers that effects the costing of disassembly of products which will be relevant information for the current design disassembly process or even for future designs.

ABC is chosen because of the noted superiority in cost-tracing, separation of direct and indirect costs, higher accuracy, and its capability to blend into the Activity-Based Management (ABM) systems that more and more companies are employing (Turney, 1991). A motivating example for its use in an environmental context can be found in (Brooks, 1993) where it is described how Activity-Based Costing and environmental aspects can be combined to give companies the ability to identify more accurately

those plants and products which are driving up their environmental expenditures.

The focus of ABC is on accurate information about the true cost of the product. This system is based on the premise that products consume activities, activities consume resources and resources consume cost (Kaplan, 1984). In ABC the assumptions is made that products or services do not directly use up resources but consume activities. Hence in ABC, the cost of a product equals the sum of the costs of all activities that must be performed in the realization of the product (R.Cooper, 1990).

The goals of ABC can be achieved by managing the activities. It is a process of relentless and continuous improvement of all aspects of business. This involves a continual search for opportunities to improve which in turn involves a careful and methodical study. ABC assumes that activities consume resources and products consume activities. It uses two –stage procedure to calculate the product costs: it traces resources costs to the activities, and then traces the cost of activities to the products.

Once the activities identified then the cost of the product is calculated based on the consumptions of activities. In the process, cost drivers will be determined and information on the key elements of these drivers will be beneficial for process improvements and cost reduction (A.Gunasekaran, 1999).

Basically, ABC is a two-stage approach for allocating indirect costs to product units based on cost drivers at various levels of activity. In the first stage,

resource costs are assigned to cost pools which correspond to the various types of activities performed by the organization.

Each type of activity is then costed out individually based on the total costs of resources consumed divided by the volume of activity performed. In second step, activity costs are assigned to the products, customers and services that benefit from or create the demand for the activities (T.Raz and D.Elnathan, 1999).

The activities in each level are identified and matched with the level bases used to assign cost to products. Next, cost drivers are selected for each activity. The cost associated with each activity is added and cost per allocation-base unit is calculated. The cost of a product is then calculated as the sum of costs of all activities which were included in its production at various levels (T.Raz and D.Elnathan, 1999).

1.2 Design for Disassembly (DfD)

Design for disassembly (DfD) is another aspect of design engineering where leads to the correct identification of design specifications to minimize the complexity of product structure (Gungor and Gupta, 1999). This can be done by minimizing the number of different parts, increasing the use of common materials, optimizing the alignment between various components to facilitate disassembly with risking assemblability, functionality and structural soundness of a particular product (A.Desai & A.Mital, 2003).

The important role of product design in cost management is undeniable as 70% of a final product cost is determined at the design stage (Boothroyd & Dewhurst, 1994). Cost accounting should help product development to design products which can be disassembled with the most optimized total cost while also taking account of other important design criteria.

Design for disassembly is a technique that is used when designing new products to make their disassembly for recycling easier. Disassembly is defined as the degree of easy disassembly (Mok et al., 1997). Major related gains (80-90%) of disassembly tends to be determined at the product design stage, therefore it is vital to incorporate cost factors and environmental consideration into product design (Kriwet et al., 1995). Product recovery options achievable by design for disassembly may be classified into categories such as repair, refurbish, remanufacture, recycling and cannibalization (Thierry et al., 1995).

Disassembly is an important issue to today's manufacturers. However, this becomes a problem because disassembling products are costly. There are limited mechanical processes for disassembling old products such as shredding and disassemblies are largely in manual and labor and time intense (G.Villalba, 2004). These activities urge designers to design new products that are more environmental friendly and economical. The environmental impacts could only be minimized only if the products can be disassembled easily and cost effectively.

Feldmann et al. (1999) touched upon the issue of disassembly costs. Recycling costs and benefits differ for specific fractions of recovered materials. The more

important economic considerations to be taken into account during the disassembly process include such factors as value added to products and materials during manufacturing, disassembly cost and revenue per operation and the penalty if poisonous materials are not completely removed (Ad de Ron, 1995).

Therefore the main obstacle facing future disassembly and recycling technology are more economic than technological in nature (Boks and Templemon, 1998). To attain this objective, the economic feasibility of integration between ABC and DfD become important. Designing products so that can be disassembled and recycled is an area that is of increasing importance to today's industry. Manufactures can no longer continue to encourage consumption, without beginning to take responsibility for the way in which their products will be dealt at the end their life cycle (Tracy Dowie, 1994).

1.3 Integration of ABC and DfD

An optimal disassembly strategy should optimize the use of manual labor, specialized tools and etc. This can be achieved by implementing ABC findings in DfD technique as this can be used as basis to be considered in disassembly process design of existing product or design of future product. Under this approach, the cost drivers for product disassembly such as usage of tools, usage of cleaning materials and labor hours which will be identified in the ABC analysis will help in designing future products that will be cost-conscious during disassembly.

The cost of handling, sorting and disassembly will play an ever more important role in determining the cost of disassembly. It is also vital to analyze that cost of disassembly do not exceed the value recovered in certain cases. For example, imagine that from the ABC analysis it is identified for a particular product disassembly the major cost drivers are number of screws. Therefore, for future design the designer can use this information to design product with less screw which eventually minimize cost when it has to be disassembled.

This will help to answers uncertainties associated with the information used in product cost assessments and most importantly to identify those process and product design that has the largest influence on the disassembly cost and the relationship of cost drivers and cost drivers.

The objective of this paper is to illustrate identification of critical activities and cost drivers, carry out process improvements on demanufacture activities and operations to improve cost and simulate scenarios in cost behavior due to improvements in demanufacture job design. This is made possible with the integration of DfD and ABC in various stages of the study in an actual production.

2. Case Study

The study was done in an electronics manufacturing system company located in Malaysia. This company provides

design, engineering, manufacturing and post-manufacturing services such as refurbishment and out of warranty repairs and upgrade. The product that was chosen to illustrate is a particular major runner hand phone model out of the 6 types of models that goes through refurbishment and repair as part of its agreement to repair this phone under the warranty period provided to its customers. In view of the current costing system that was practiced it was established that it is beneficial for ABC to be proposed to be able to estimate product cost accurately. This was also important to identify critical activities and cost drivers for cost improvement which would lead to a better technique in addressing demanufacture activities in this case.

Currently the cost system that is being practiced in this company is a conventional costing system or also known as traditional costing system. This costing system focuses on volume-based cost allocation as a reference for pricing a single unit cost of a particular product. The cost factors inclusive in this system are material, overhead, labor, manufacturing overhead and indirect material.

All the above cost except for material are then summed together and divided with product volume to compute the cost per unit product. This cost is added with material standard cost to generate the final product cost. It can be simplified as the Equation (1).

$$\text{Cost per unit product} = \frac{\text{Material standard cost} + \text{IDL} + \text{Direct Labor} + \text{Mfg. Overhead}}{\text{Volume}} \quad (1)$$

Basically, there are 5 parts from the hand phones that are salvaged and recycled in order to reduce the material standard cost. The items are volume keypad, back housing, front housing, keypad and antenna. The percentage that can be salvaged is around 50% of the total volume received. The process begins with the pre-screen, sorting, debug, test, assembly and finally packing. The data collection and study will be emphasized on the study of cycle time, process, material, indirect labor, direct labor, space, utilities and equipment depreciation. This is done inline with current cost structure that breaks down the pricing of product into these areas. Furthermore, each of the indirect labor involved is studied to find the job description that correlates with the particular product.

Based on the forecast and the cycle time analysis, the percentage of product and cycle time is calculated to be allocated accordingly with the related cost. With this information the identification of cost pool and cost driver is possible. Also, with further study on the equipment, space and utilities the accurate costing using ABC is captured with the respective cost drive identified. The study also involves on the analysis of value and non-value added activities in each work stations for further improvements. This will be beneficial as one of the cost drivers are cycle time and with the reduction or elimination of non-value activities it can further drive down the cost.

2.1 Methodology

The methodology for this study is illustrated in Figure 1 where it is in a cycle of Plan-Do-Check-Act. Firstly, various manufacturing costs are traced to activity centers by first stage cost assignment and subsequently to product with second stage cost drivers as per Figure 2. Each activity will be assigned to a cost driver that is required for the product. The cost drivers chosen must accurately measure the actual consumption of the activities. Cost driven will be chosen based on ease of obtaining data relating to the cost drivers. In the data collection and analysis the flow can further broken down to 3 main segments which are data collection method, collected data and data analysis.

The main data collection methods are by discussion, interview, line study or observation and document reference. Meanwhile the collected data consists of process flow, time study, work instructions, bill of material, current cost price, cost driver, cost pool, value and non-value add activities, equipment pricing, head counts, job descriptions, labor wages, forecast, space and utilities cost. From these data, the analysis that were performed are on process flow, time study, value and non-value added activities, forecast, bill of material and layout.

Next, the product cost is calculated based on ABC system to be compared with the traditional costing system cost. This is to highlight the difference between these 2 costing system and consequently DfD techniques and process improvement is applied to the changes in work design and work practice to reduce the cost driver activities on various stations based on the cost driver findings. DFD techniques are put into action for process improvement which should lead to cost reduction or higher capacity utilization.

With the implementation or simulation on the improvements, a new ABC cost is recalculated to show the new product cost and possible savings that can be achieved. Based on the result of the study, various scenarios will be tested to create a platform for future ABC cost system for products in the same range.

3. Results

The first part of the analysis was concentrated on calculating the cost price of the product based on ABC system to show the difference with traditional costing. The result is illustrated in Table 2 where it shows that there is a difference of USD\$5,932 per month for the monthly volume produced by the company.

Based on the cost pool percentage comparison for all the models, the weightage for each of the resources is determined and it is used to build a resources verses cost drivers matrix as shown in Table 3 & 4. The ratings are based on the multiplication of the weightage and relation of the cost driver with the respective resources. The relation of the cost driver and the resources is determined from the discussion and interview with the employees involved from various departments and also the study on all of the activities.

From this matrix, a cost driver weightage table and Pareto chart as Table 5 and Figure 3 is built to determine the cost drivers that contribute 80% to the cost which in this case are scraps, defects, process and volume. These cost drivers is studied further to improve and reduce the overall product cost.

Once the major cost drivers are identified, it is then analyzed and improved to reduce the resources. In this case, the improvements can be categorized into 2 main categories, which are process improvements and scrap and defect improvements.

For process improvements, it is concentrated on increasing value added activities and reducing non-value added activities. Plus, line balancing is performed in the production line to eliminate waste in any particular stations. One way of doing this is by allocating equal work for each station by using the element process time as guidance.

Basically, the process are combined based on the each individual task time to make each stations total time to be as close as possible with the maximum time allowable for one product which is called TAKT time. This TAKT time is determined based on the volume, working days and working hours. Furthermore, process time is also improved by simplifying the working method or by introduction of fixtures or jigs. The summary on the process improvements done are illustrated as per Table 6.

This shows by improving the process the cycle time, non-value added percentage, headcount and balancing loss has been reduced meanwhile the value added percentage has been increased. This will translate into savings and a new reduced cost price for the product as process time and headcount are among the cost drivers that contribute to the cost of the product.

Next, the improvements were performed in improving scrap and defect. Current percentage of material that is reused from the sorting station averages at 50%. These materials are salvaged during the sorting process by the operators. The major parts that are involved are antenna, volume button, keypad and housing. Actions were taken to solve the issues after further discussion with the QA team on the acceptance criteria. Sample of visual check were placed on the line to assist the sorting operators to properly identify the reject criteria's.

Once all the actions were implemented on the line, the result was monitored for 2 weeks and the defect or scrap rate was monitored. The result shows a reduction of 11% in the materials that are being scraped. From this, the material consumption can be reduced to only 40% from previous of 50%. Indirectly, this contributes to cost saving as lesser material need to be purchased for the refurbishment of the products. When the improvement is completed, ABC and traditional costing is used again to calculate the new product cost to see the price difference before and after improvement. The result can be summarized as per Table 7.

4. Conclusion

The result of this study clearly shows that there is cost difference between ABC and traditional costing system. For some product, the ABC cost is much lower than the traditional costing and vice versa. Next, the critical activities and cost drivers present in the demanufacturing of the product has been identified. The ABC cost system implementation provided the product cost and system information such as the cost drivers regarding to the activities or processes to be used as guidance for improvement.

From ABC cost simulations on various products of the same nature it was found that traditional costing doesn't necessarily over quotes products but it also under quotes at times. This shows that ABC cost system provides information or cost to remain competitive in this business.

By identifying cost drivers, improvements was performed and it showed the comparison between the cost based on traditional cost system and ABC cost system before and after improvement. The final ABC cost system price is USD 26.27 compared to the initial cost price of USD 27.67 based on the traditional costing.

Improvements were applied to cost drivers that drivers 80% of the product cost based on the Pareto analysis of the cost drivers relationship and resources. The ABC system has helped to identify the area that has to be improved to see significant cost improvement. Line balancing, process improvement and work method improvement was performed to these cost drives and this significantly reduced the product price.

This proves that ABC cost system can be combined with other analysis or engineering tools for improvements on the costing and it is applicable for the demanufacturing environment as proven by this study. Furthermore, the changes of cost shows that the cost structure is reactive with cost drivers improvements.

However, there are many hurdles in applying ABC cost system in a manufacturing environment as faced in this study. The data collection is one of the major hurdles in using this system as not that it involves only on confidential data such as material price or employee's salary but it also touches sensitive issues such as detail tasks of the employees in his or her daily job.

Furthermore, the required informations are many and some data's are not able to be retrieved readily from databases. Some examples are such as correlation of cost driver and resources or even the percentage of employee's tasks. These type data has to be studied carefully and it is time consuming. At times, estimation is made where the information is not available based from observation or discussion with related parties.

As conclusion, the implementation of ABC cost system in a demanufacturing environment is able to determine the cost of product as accurate as possible. Furthermore, it helps to determine the cost drivers for each resource that drives the cost of a product which consequently helps to narrow down the scope for

improvement. When combined with other manufacturing or engineering tools such as line balancing, value analysis or process improvement it becomes a powerful tool for cost reduction of product.

Based on the experience and successfulness in conducting this study, it can be said that in future one of the possible approaches is to further study implementation of ABC cost system in various electronics demanufacturing environment. This is to see the interrelation of the cost drivers in these environments. From the findings, a table or guide can be done to predict the cost drivers of a particular industry as this will be beneficial for cost improvement target.

In future the findings of this study should be used at design stage of a product that will go through disassembly as this will help to reduce cost and ease the process during disassembly. The cost drivers for disassembly should be improved as much as possible during the initial design stage as this will play a significance role in the cost of product during disassembly.

When this both future undertaking is combined, it will be able to design an ABC interface that will be able to predict and reduce cost price, cost driver and other related aspects for any product in the same environment.

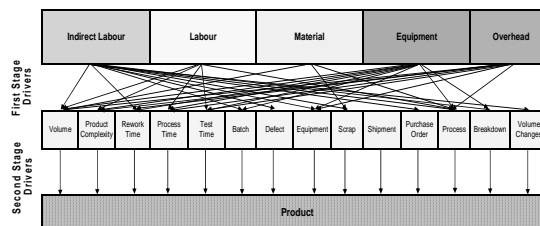


Figure 2 : ABC structure on the first & second stage drivers

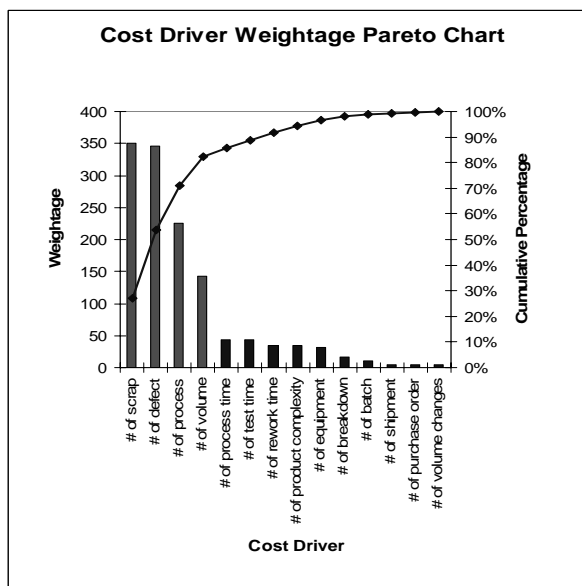


Figure 3 : Pareto Chart to determine the major cost driver

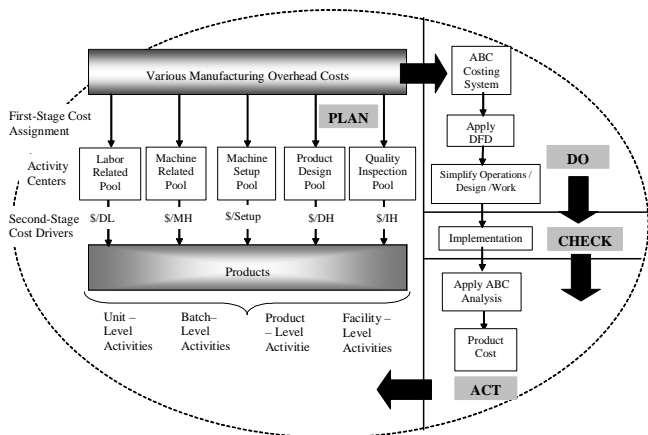


Figure 1 : Methodology in the ABC & DfD implementation for the case study

Table 1: Value and Non-Value Added Activities in Product Cycle Time

Process	Total Time (sec)	Value Add		Non-Value Add	
		Time (sec)	Percentage (%)	Time (sec)	Percentage (%)
Prescreen	197	177	90%	20	10%
Sorting	261	89	34%	172	66%
Assembly	811	741	91%	70	9%
Debug	980	968	99%	12	1%
Test	240	191	80%	49	20%

Table 2 : Total product cost price comparisons for ABC and traditional costing

Product	Costing System	IDL	Material	Equipment	Labor	Overhead	Total	Cost Delta / Month
Handphone	ABC	\$1.00	\$24.35	\$1.04	\$0.79	\$0.24	\$27.42	
	Traditional Costing	\$1.01	\$24.35	\$1.16	\$0.88	\$0.27	\$27.67	
	Delta	\$0.012	\$0.000	\$0.118	\$0.090	\$0.028	\$0.247	\$5.932

Table 3 : Resources verses cost drivers matrix

Resources	Weightage	Cost Driver															
		# of volume	Rating	# of product complexity	Rating	# of rework time	Rating	# of process time	Rating	# of test time	Rating	# of batch	Rating	# of defect	Rating		
Indirect Labor	5	3	15	0	0	2	10	0	0	0	0	1	5	2	10		
Labor	4	4	16	5	20	0	0	4	16	4	16	0	0	0	0		
Material	84	1	84	0	0	0	0	0	0	0	0	0	0	4	336		
Equipment	6	4	24	2	12	4	24	4	24	4	24	1	6	0	0		
Overhead	1	4	4	2	2	1	1	3	3	3	3	0	0	0	0		
Total	100	16	143	9	34	7	35	11	43	11	43	2	11	6	346		

Table 4 : Resources verses cost drivers matrix

Resources	Weightage	Cost Driver															
		# of equipment	Rating	# of scrap	Rating	# of shipment	Rating	# of purchase order	Rating	# of process	Rating	# of breakdown	Rating	# of volume changes	Rating		
Indirect Labor	5	1	5	3	15	1	5	1	5	4	20	1	5	1	5		
Labor	4	0	0	0	0	0	0	0	0	4	16	0	0	0	0		
Material	84	0	0	4	336	0	0	0	0	2	168	0	0	0	0		
Equipment	6	4	24	0	0	0	0	0	0	3	18	2	12	0	0		
Overhead	1	3	3	0	0	0	0	0	0	3	3	0	0	0	0		
Total	100	8	32	7	351	1	5	1	5	16	225	3	17	1	5		

Table 5 : Cost driver weightage for Pareto chart

Cost Driver	Weightage	%	Cumulative
# of scrap	351	27.1%	27.1%
# of defect	346	26.7%	53.8%
# of process	225	17.4%	71.2%
# of volume	143	11.0%	82.2%
# of process time	43	3.3%	85.6%
# of test time	43	3.3%	88.9%
# of rework time	35	2.7%	91.6%
# of product complexity	34	2.6%	94.2%
# of equipment	32	2.5%	96.7%
# of breakdown	17	1.3%	98.0%
# of batch	11	0.8%	98.8%
# of shipment	5	0.4%	99.2%
# of purchase order	5	0.4%	99.6%
# of volume changes	5	0.4%	100.0%

Table 6 : Summary of results before and after process improvement for product

	Before Improvement	After Improvement	Delta
Cycle Time (sec)			
Prescreen	197	189	8
Sorting	261	241	20
Assembly	811	802	9
Debug	980	980	0
Test	240	240	0
Total Reduction	2489	2452	37
Value Added (%)			
Prescreen	90%	94%	4%
Sorting	34%	37%	3%
Assembly	91%	91%	0%
Debug	99%	99%	0%
Test	80%	80%	0%
Total Increase	394%	400%	7%
Non - Value Added (%)			
Prescreen	10%	6%	-4%
Sorting	66%	63%	-3%
Assembly	9%	9%	0%
Debug	1%	1%	0%
Test	20%	20%	0%
Total Reduction	106%	100%	-7%
Headcount	84	77	7
Line Balancing Loss	33%	14%	20%

Table 7 : ABC Cost per unit (USD) comparison.

COST	ABC Improved Price	ABC Previous Price	Traditional Pricing Improved	Traditional Pricing Previous
IDL	\$ 1.00	\$ 1.00	\$ 1.01	\$ 1.01
MATERIAL	\$ 23.29	\$ 24.35	\$ 23.29	\$ 24.35
EQUIPMENT DEPRECIATION	\$ 1.03	\$ 1.04	\$ 1.16	\$ 1.16
LABOR	\$ 0.70	\$ 0.79	\$ 0.79	\$ 0.88
OVERHEAD	\$ 0.24	\$ 0.24	\$ 0.27	\$ 0.27
TOTAL	\$ 26.27	\$ 27.42	\$ 26.52	\$ 27.67
DELTA	\$ 1.15		\$ 1.15	

References

[1] Anoop Desai, Anil Mital*, 2003, Evaluation of disassemblability to enable design for disassembly in mass production, International Journal of Industrial Ergonomics 32 (2003) 265–281.

[2] A.Gunasekaran, Design of Activity Based Costing in a small company, 1999, Computers and Industrial Engineering 37 (1999) 413-416.

[3] Bert Bras and Jan Emblemvåg, 1995, the use of activity-based costing, uncertainty, and disassembly action charts in demanufacture cost assessments ,1995 ASME Advances in Design Automation Conference.

[4] Brooks, P. L., Davidson, L. J. and Palamides, J. H., “Environmental compliance: You better know your ABC’s”, Occupational Hazards, February (1993), pp. 41-46.

[5] Boks, C., Templemon, E., 1998. Future disassembly and recycling technology. Results of a Delphi study. Futures 30 (5), 425–442.

[6] Boothroyd.G & Dewhurst.P, 1994, Product design for manufacture & assembly, 2-4.

[7] de Ron, A., Penev, K., 1995. Disassembly and recycling of electronic consumer products: an overview. Technovation 15 (6), 363–374.

[8] Feldmann, K., Traunter, S., Meedt, O., 1999. Innovative disassembly strategies based on flexible partial destructive tools 159–164..

[9] Kriwet, A., Zussman, E., Seliger, G., 1995. Systematic integration of design-for-recycling into product design. International Journal of Production Economics 38, 15–22.

[10] Mok, H.S., Kim, H.J., Moon, K.S., 1997. Disassemblability of mechanical parts in automobiles for recycling. Computers and Industrial Eng. 33 (3–4), 621–624.

[11] Thierry, M., Salomon, M., Van Nunen, J., Van Wassenhove, 1995. Strategic issues in product recovery management. California Management Review 37 (2), 114–135.

[12] Tracy Dowie, 1994, Green Design, World Class Design to Manufacture, Vol. 1 No. 4, 1994, pp. 32-38.

[13] Turney, P. B. B., “How Activity-Based Costing Helps Reduce Cost”, Journal of Cost Management for the Manufacturing Industry, Vol. 4, No. 4 (1991), pp. 29-35.

Condensation of Refrigerant R-22 in the Annulus of Horizontal Double-Tube Condenser with an Enhanced Inner Tube

R. Tiruselvam¹, Mohd. Zainal Yusof¹, Vijay R. Raghavan^{1*}

¹Fakulti Kejuruteraan Mekanikal & Pembuatan
Kojel Universiti Teknologi Tun Hussein Onn, Parit Raja
Tel: +60-125698389, E-mail: tiruselvam_ramahlingam@yahoo.com

Abstract

In this study, condensation heat transfer characteristics of an enhanced tube are studied experimentally in condensation of refrigerant R-22. Heat transfer and pressure drop correlations for the enhanced double tube condenser are obtained that can aid in development non-CFC refrigerant systems. Accurate values of condensation heat transfer coefficient and convective heat transfer coefficient are required to obtain correlations suitable for design purposes of an augmented helical concentric pipe heat exchanger (double-tube). The condensation coefficients are determined using the Wilson Plot technique. The experimental apparatus consists of forced circulation loops of refrigerant and cooling water. The superheated vapour flows from the compressor to the annulus of a horizontal double-tube condenser and is condensed to sub-cooled liquid in counter flow pattern. From experimental data and applying analytical equations condensation overall heat transfer coefficients are obtained at constant heat flux. The results of the augmented condensation study are presented graphically and compared with the non-augmented case.

Key words:

Refrigerant, Condensation Enhancement, Tube Annulus, Wilson Plot Technique

1 Introduction

For the past few decades, CFCs have been extensively used in refrigeration and air-conditioning field due to their excellent thermodynamic and chemical characteristics. Due to the Montreal protocol signed in 1987 and with subsequent amendments, however, CFCs were completely phased out as of January 1996 in developed countries while they are allowed for further use for the grace period of 10 years in developing countries [1]. The refrigeration and air-conditioning industry has grown significantly over the past few decades and will continue to do so all over the world for the coming years. Thus, it is necessary to develop more energy efficient refrigeration and air-conditioning equipment. It has been shown by Jung et al. [2] that refrigerant R22 gives a higher heat transfer coefficient than of R410A and of R407C on an enhanced tube. For all refrigeration equipment, heat exchangers are the basic components. Their efficiency should be increased for overall improvement of the system efficiency. In this study, condensation heat transfer characteristic of enhanced tubes, which are currently used in the design of double-tube condenser for the Water Source Heat Pump (WSHP) unit are studied experimentally.

*author for correspondence: vijay@kuittho.edu.my

1.1 Scope of the Study

This study is conducted on water source heat pump unit (Figure 1) with 19,500 Btu/hr (5.72 kW) cooling capacity for cooling mode. The condensation heat transfer coefficient developed is for units using refrigerant R-22 as a refrigeration medium and water as cooling medium (coolant).



Figure 1: (a) Water Source Heat Pump (b) Condenser Coil

The scope of this study is:

1. Wilson Plot Method - This study utilizes the condensation correlation available from previous research. With this as the starting point, the correlation obtained from the current test will be

substituted to arrive at a correlation based on the experiment conducted.

2. Straight Length Prototype Unit - A prototype unit of the WSHP comprising Double Tube Heat Exchanger (Condenser) System is used for the test. The Wilson plot method applies to straight horizontal length of condenser tubes due to the fact that correlations for the fully developed turbulent flow currently available are for straight length tubes. This prototype straight length unit is the initial test section. The data and results obtained from this experiment are to be used later for coiled Double Tube Heat Exchanger, taking into consideration of the secondary flow created by the tube coiling effect. See Figure 2.

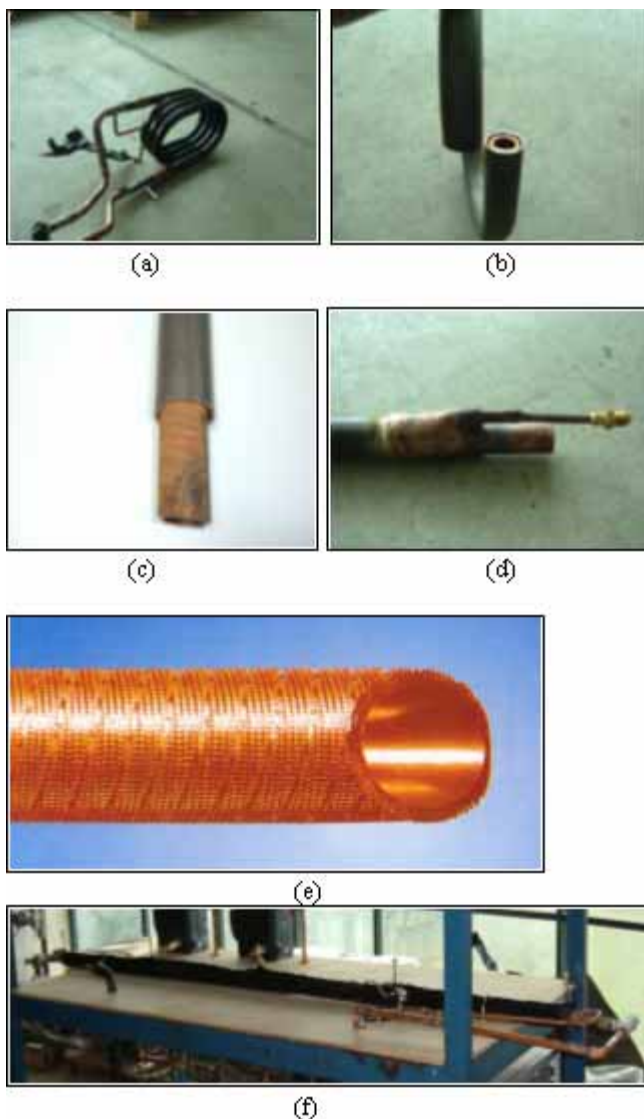


Figure 2: (a) Double Tube Heat Exchanger (b) Cross-section View (c) Tube Configuration (d) Inlet Section (d) Turbo-C Inner Copper Tube (f) Straight Length Test Section

1.2 Methodology

The separation of individual heat transfer resistances from the overall heat transfer resistance of the system is extremely important in obtaining heat transfer correlation for the annulus side and tube side. Accurate values of condensation heat transfer coefficient and convective heat transfer coefficient are required to obtain correlations suitable for design purposes of an augmented helical concentric pipe heat exchanger (double-tube). Local heat transfer coefficients can be found by direct measurement of the temperature drop across the convective film and the condensate film by using thermocouples located in the wall of the tube and in the bulk stream.

This method works well for a single-resistance heated tube for which thermocouples can be attached to the tube wall without disturbing the fluid flow in the vicinity of the thermocouple. In the design of heat transfer equipment, overall convective or film condensation coefficients are usually preferred to local coefficients. Although local coefficients can be suitably integrated over the tube length to give the average coefficient for the whole tube, this cannot be conveniently done for many systems and configurations. As for the double-tube heat exchanger, with coolant water in the inner tube and condensing refrigerant in the annulus, such an arrangement makes it impossible to put a sufficient number of thermocouples inside the shell (annulus) without seriously affecting the refrigerant flow in the annulus. Thus, a different approach is required for such situations, hence the Wilson Plot Technique.

2 Literature Review

2.1 Wilson Plot Method

A useful technique for determining individual resistance from an overall resistance was devised by Wilson (Cited by Briggs et al [3]) in 1915. Wilson was interested in determining the effects of water temperature and velocity on the overall coefficient for a steam condenser. His technique was modified by Young and Wall (Cited by Briggs et al [3]) in 1957 to obtain shell and tube side heat transfer correlation for concentric pipe heat exchanger. Wilson Plot method was originally developed as a design aid but gained prominence as a graphical tool to determine internal heat transfer correlations for heat exchanger performance test. The coupling between performance and the internal heat transfer rate is an important factor in determining the flexibility of the proposed analytical modeling of double-tube condenser. This argument is especially true if an overall heat balance is used to represent the overall heat transfer process within the design condition of a heat exchanger. The Wilson plot method used in this research work is to demonstrate the flexibility and scope of the proposed experiments and therefore illustrate the usefulness of a proposed experimental design. This method also demonstrates the use of Wilson

Plot to determine the internal heat transfer coefficient as a function of the internal flow conditions.

2.2 Description Of Methodology

In most heat exchanger condensing processes, the primary measurements taken connect the energy transferred to the temperature difference between two fluid streams. The overall rate of heat transfer can be calculated from:-

$$Q = U A_o \Delta T \tag{1}$$

The objective of the condensation test is directed at determining the condensate heat transfer resistance from the overall thermal balance. When a vapour condenses on the outside of a tube which is internally cooled by flowing liquid (Fig. 5), the component thermal resistances are coupled to the overall resistance by:-

$$\frac{1}{A_o U} = R_o + R_w + R_i = \frac{1}{A_o h_o} + R_w + \frac{1}{A_i h_i} \tag{2}$$

It is noted that there will be some heat loss through the outer steel tube as the outer surface of the condenser coil will not be insulated during usage in actual unit application. This heat loss is an added advantage to the entire heat rejection process by the condenser coil. Since the thermal conductivity of steel is less compared to the thermal conductivity of copper ($k_{steel} \ll k_{copper}$), the outer surface is assumed to be adiabatic. Additional precaution is taken by insulating the outer steel pipe with 1" thick Superloon insulation material to create a adiabatic outer surface. Equation (2) can then be transformed to:-

$$\frac{1}{U} = \frac{A_o}{A_i} \frac{1}{h_i} + \left[A_o R_w + \frac{1}{h_o} \right] \tag{3}$$

The interpretation of equation (3), which has the linear form $y = mx + b$, if R_w and h_o are constant, is the basis of the Wilson Plot method. Reviews on the use of this method in different types of heat exchangers have been given by Shah (cited by Deans et al [3]), where:

$$y = \frac{1}{U} \tag{4}$$

$$m = [C_i]^{-1} \tag{5}$$

$$x = \left\{ \left(\frac{A_o}{A_i} \right) \frac{1}{h_i} \right\} \tag{6}$$

$$b = \left\{ A_o R_w + \frac{1}{h_o} \right\} \tag{7}$$

A linear regression of the function y above on x gives the least-square deviation values of m and b . These values of m and b can be later converted to the constants value for the outer and inner (C_o and C_i) heat transfer coefficient.

Equation (7) can then be used to generate a straight line graph which describes the overall heat transfer process across the condenser tube when the coolant temperature and mass flow rates change but quality of heat transfer is constant. As consequence of this, the internal temperature difference and heat transfer coefficient are balanced at different values so that while the overall internal heat transfer coefficient varies, the overall heat transfer and external heat transfer coefficient remain unchanged. Given such a test series, a line can be plotted as shown in Fig. 3.

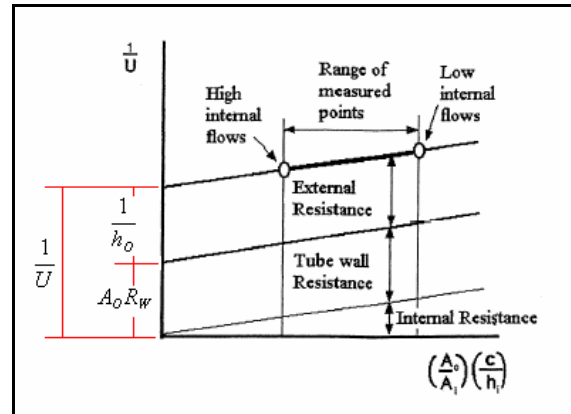


Figure 3: The Wilson Plot- General features

The experimental points extend from a high internal flow rate, with low overall heat transfer resistance, to a low internal flow rate and high overall heat transfer resistance. A line drawn through these points intersects the U^{-1} – axis at a point where the internal heat transfer resistance approaches zero, and thus the overall heat transfer resistance at this point is simply the sum of wall resistance and the external heat transfer resistance. The internal heat transfer coefficient, at any flow rate, can then be determined independently using equation (3). It should be noted that the accuracy and reliability of the Wilson Plot Method depends on the validity of the correlation used to describe the internal and external heat transfer coefficient.

2.3 Development of Correlation

Since the analysis of the turbulent flow condition in the inner tube of the condenser will be used to determine the heat transfer and internal heat transfer coefficient, greater emphasis is placed on determining empirical correlations to be used here. A classic expression for computing the Nusselt's number for fully developed (hydro-dynamically and thermally) turbulent flow in a smooth circular tube is given by Dittus-Boelter equation (8) and Sieder-Tate equation (9).

$$Nu_D = 0.023 (Re_c^{0.8} Pr_c^{0.4}) \tag{8}$$

$$Nu_D = 0.027 \left(Re_c^{0.8} Pr_c^{1/3} \right) \left(\frac{\mu_c}{\mu_{cw}} \right)^{0.14} \quad (9)$$

Although equation (8) and (9) are easily applied and are certainly satisfactory for theoretical application, errors as large as 25% may result from their use. Such errors may be reduced to less than 8% through the use of more recent, but more complex, correlation [6]. One such correlation, which is widely used is attributed to Petukhov (cited by Bhatti et al [6]), is of the form

$$Nu_D = \frac{(\frac{f}{8}) Re_D Pr}{1.07 + \left(\frac{900}{Re_D} \right) - \left(\frac{0.63}{1 + 10 Pr} \right) + 12.7 (\frac{f}{8})^{1/2} (Pr^{1/4} - 1)} \quad (10)$$

By converting this Petukhov equation to the form of overall internal heat transfer coefficient,

$$\bar{h}_i = \left(\frac{k_c}{d_o} \right) \left(\frac{(\frac{f}{8}) Re_D Pr}{1.07 + \left(\frac{900}{Re_D} \right) - \left(\frac{0.63}{1 + 10 Pr} \right) + 12.7 (\frac{f}{8})^{1/2} (Pr^{1/4} - 1)} \right) \quad (11)$$

Heat transferred from the refrigerant to the cooling water during condensation can be obtained from,

$$Q = \dot{m}_c C_p (T_{O,C} - T_{i,c}) \quad (12)$$

where

$$U = \frac{Q}{A_i \Delta T_{(LMTD)}} \quad (13)$$

$$\Delta T_{(LMTD)} = \frac{(T_s - T_{O,C}) - (T_s - T_{i,c})}{\ln \left(\frac{T_s - T_{O,C}}{T_s - T_{i,c}} \right)} \quad (14)$$

$$R_w = \frac{\ln(d_o/d_i)}{2\pi k_c L} \quad (15)$$

By rearranging equation (3),

$$\bar{h}_o = \left[\left(\frac{1}{U} \right) - (A_o R_w) - \left(\frac{A_o}{A_i \bar{h}_i} \right) \right]^{-1} \quad (16)$$

or;

$$\bar{h}_o = \left[\left(\frac{A_i \Delta T_{(LMTD)}}{Q} \right) - (A_o R_w) - \left(\frac{A_o}{A_i \bar{h}_i} \right) \right]^{-1} \quad (17)$$

The use of equation (11) in equation (17) will give a simple approach to the Wilson plot technique. The basic method of maintaining the condition of one fluid and changing the other still applies here.

3. Experimental Set-up and Procedure

3.1 Instrumentation

The experimental apparatus, shown schematically in Fig. 4, consists of forced circulation loops of refrigerant and cooling water. The superheated vapour is generated from the compressor flows to the horizontal test section of the double-tube condenser. The vapour is condensed from superheated vapour to sub-cooled liquid. The refrigerant flows in the annulus of the double-tube with counter flow pattern to the cooling water in the inner tube. The condensed

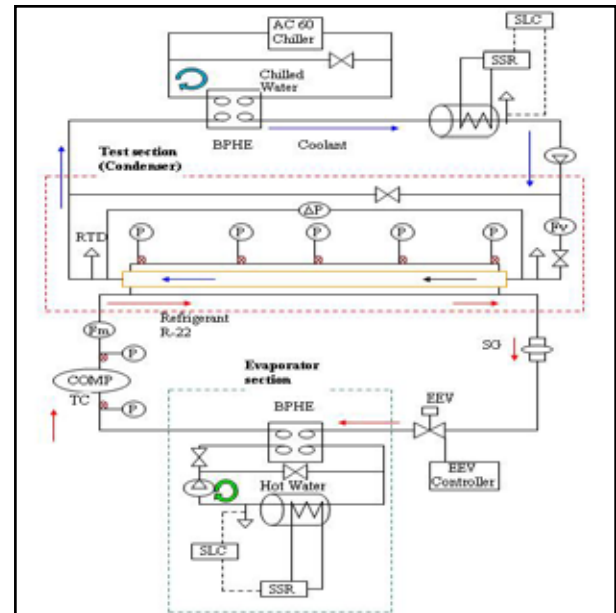


Figure 4: Schematic Diagram of Experiment Setup

refrigerant then flows to the braze plate heat exchanger via the mass flow meter (where the mass flow of the refrigerant is measured) and expansion valve (expansion valve is used to meter the flow of liquid refrigerant entering the evaporator at a rate that matches the amount of refrigerant being boiled off in the evaporator). The refrigerant is then evaporated in the Braze plate heat exchanger and flows back to the compressor as 100% vapour. The cooling water is pumped into the test section via the volume flow meter by the water pump. After exchanging heat with the refrigerant in the test section, the cooling water flows to the condenser side braze plate heat exchanger (BPHE unit used to reject heat from the water and maintain the cooling water temperature). The heat rejection rate of the BPHE is controlled by the flow rate of chilled water into the BPHE. Pressure and temperature data is collected from the experiment by pressure transmitters (pressure reading) and thermocouple (temperature reading) for the test section and the inlet and outlet of the compressor. These data is collected by the computer via the data recorder and data logger (DAQSTATION DX220 Communication Interface).

3.2 Testing Procedure

The purpose of the ANSI/ASHRAE Standard 37-2005 (Method of Testing for Rating Electrically Driven Unitary Air-Conditioning and Heat Pump Equipment) [4] is to provide test methods for determining the cooling capacity of unitary air-conditioning equipment and the cooling or heating capacities, or both, of unitary heat pump equipment. These standard do not specify methods of establishing rating which involve double tube condenser but is proves as an initiation point for our test setup. Further procedures and installation methods are adapted to provide the relevant steps required for the Wilson Plot technique procedures.

4. Analysis Of Test Result

The results of the condensation experiment conducted to obtain the heat transfer coefficient for double tube heat exchanger is discussed below.

4.1 Internal Heat Transfer Coefficient.

The internal heat transfer coefficient of the condenser water flow in the inner copper tube (h_i) is shown in Figure 5 as a function of Reynolds Number (Re). The measured values of the h_i from the experiment is compared with the Nusselt type equation given by Petukhov (cited by Bhatti et al [6]) in equation (11). It is noted that the measured value is in close approximation with the equation, this is due to the low error percentage of the Petukhov equation and the flow configuration of horizontal circular smooth pipe used in the experiment.

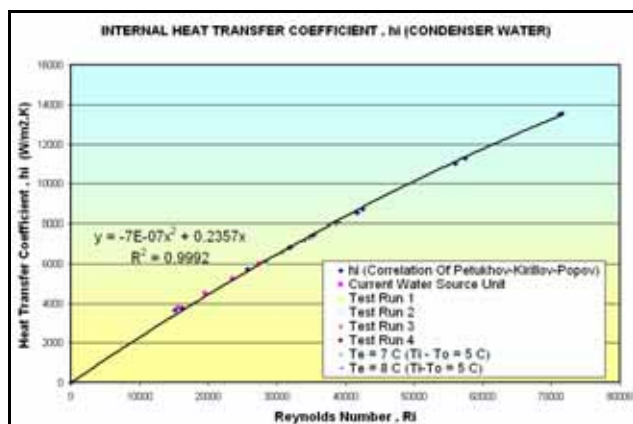


Figure 5: Internal Heat Transfer Coefficient of Water in Inner Copper Tube

4.2 External Condensation Heat transfer Coefficient

The results of the obtained heat transfer coefficient for the condensation of refrigerant R22 on the enhanced surface of the inner copper tube is illustrated in Figure 6 as a

function of measured refrigerant mass flow. The control parameter of the experiment conducted was a result of design condition of actual heat pump units;

- Temperature rise across the condenser to be maintained at 5 C.
- Refrigerant subcooling required for a range of 0-20 C.
- Suction superheat required to be constant at 3-5 C.
- Evaporation temperature from 4-8 C

With each evaporating temperature, it is found that the heat transfer coefficient, h_o , decreases with the restriction of the refrigerant flow (positioning of the EXV Valve) with a rise in internal heat transfer coefficient, h_i , and heat flux. This phenomenon can be explained as the increase in heat flux causes more vapor to condensate on the test-section tube surface. The higher condensation rate on the tube surface generates a thicker condensate layer, which in turn increases vapor to the tube wall temperature difference, ΔT_f . This results in the decrease of heat transfer coefficient, h_o , with the rise of heat flux. This ultimately reduces the flow rate of the R22. Same can be said for the other evaporating temperature where the h_o decreases with the increase of evaporating temperature.

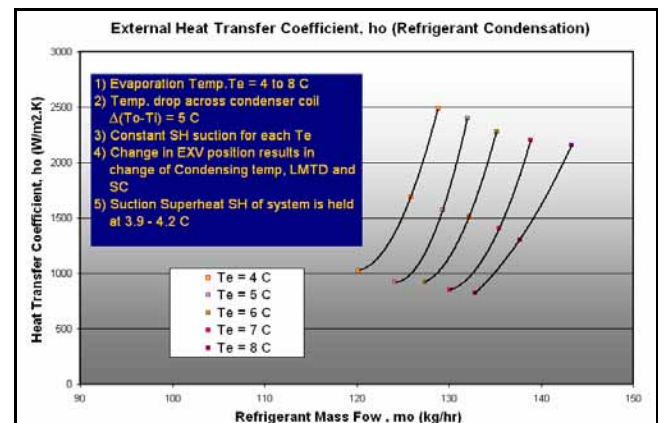


Figure 6: Condensation Heat Transfer Coefficient of R22 on the Enhanced Surface of Inner Turbo-C Copper Pipe

When the temperature rise across the condenser coil for the cooling water is fixed at 5 °C, the change in the water flow rate, refrigerant flow rate, evaporation temperature, liquid subcooling and the compressor superheat eventually gives a change in the condensing pressure. This change in the condensing pressure or the condensing saturation temperature T_{sat} , changes the $\Delta T_{(LMTD)}$ of the condensation process for each test run. Figure 7 shows the $1/h_o$ plotted against the $\Delta T_{(LMTD)}$. The results of the experiment data line shown close scatter point for a linear function $\Delta T_{(LMTD)}$ of 12 to 20 °C and slight larger scatter points for temperature above 20 °C. Further testing is required for a wider range of condensation pressure in order to develop a generalized correlation for the heat transfer coefficient, h_o , as a function of the $\Delta T_{(LMTD)}$.

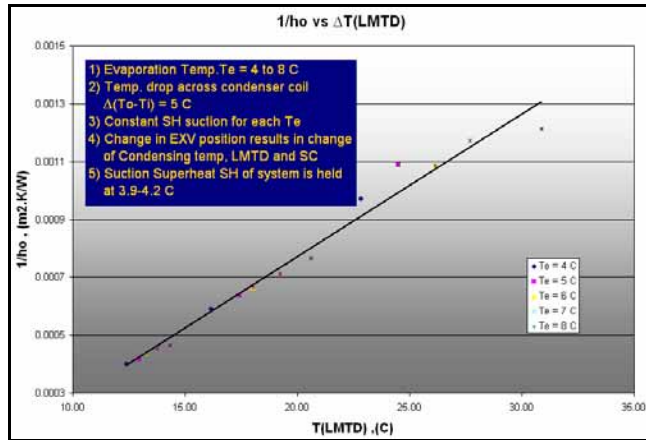


Figure 7: $1/h_o$ as a function of $\Delta T_{(LMTD)}$.

5. Conclusions

In this study, heat transfer coefficient for the cooling water flow and condensation of R22 in double tube condenser pipe is measured. All measurements are taken for the evaporation temperature of 4-8 °C, liquid subcooling of 0-20 °C, and temperature rise in cooling water of 5 °C for a fixed compressor suction superheat of 4 °C. Based upon the test results the following conclusion can be drawn.

1. The internal heat transfer coefficient, h_i , measured agrees well with the Nusselt model given by Petukhov. This is only true for the straight length pipe where for a coiled tube a different form should be used.
2. Condensation heat transfer coefficient, h_o , measured is with compliance with the design conductions of a actual water source heat pump unit. Further test of different compressor capacity will give a wider results and more generalized correlation.
3. The linear function of the $1/h_o$ and $\Delta T_{(LMTD)}$ shows a good result for this range of $\Delta T_{(LMTD)}$, but a wider range of data should be collected.

All the improvement and changes to the results in hand are currently in progress as this study is still evolving.

Reference

- [1] United Nations Environment Program. Montreal protocol on substance that deplete the ozone layer. 1989, [Final Act]
- [2] Briggs, D.E., Young, E.H., Modified Wilson Plot Techniques for Obtaining Heat Transfer Correlations for Shell and Tube Heat Exchangers, *Chemical Engineering Progress Symposium Series*, No. 92, vol. 65, 35-45, 1969
- [3] Deans, J., Sinderman, A., Morrison, J.N.A., Use of the

Wilson Plot Method to Design and Commission a Condensation Heat Transfer Test Facility, *Two-Phase Flow Modelling and Experimentation*, Edizioni ETS, 1999

- [4] ANSI/ASHRAE Standard 37-2005, Method of Testing for Rating Electrically Driven Unitary Air-Conditioning and Heat Pump Equipment
- [5] Honda, H., Nozu, S., Matsuoka, Y., Aoyama, T., and Nakata, H., Condensation of Refrigerant R11 and R113 in the Annuli of Horizontal Double-Tube Condensers with an Enhanced Inner Tube, *Exp. Thermal Fluid Sci.* 2, 173-182, 1989.
- [6] Bhatti, M.S., Shah, R.K., Handbook of Single-Phase Convective Heat Transfer, Chap. 4, Wiley-Interscience, New York, 1987

Acknowledgments

The authors would like to acknowledge with appreciation, the help and support from OYL R&D CENTER, in providing all facilities for this research.

Nomenclature

A	area (m ²)
c	Coefficient of the heat transfer correlation
d	diameter (m)
h	convective heat transfer coefficient (W/m ² K)
h_{fg}	latent heat of vaporization (J/kg)
k	thermal conductivity (W/mK)
Pr	Prandtl Number
\dot{Q}	rate of heat transfer (W)
R	thermal resistance (K/W)
Re	Reynolds Number
T	temperature (K)
U	overall heat transfer coefficient (W/m ² K)
ΔT	change in temperature (K)
ρ	density (kg/m ³)
μ	viscosity (kg/ms)

Subscripts:

c	coolant
cw	coolant wall
i	internal
l	liquid
o	outside
s	tube surface
sat	saturation
w	wall

Studies on Modelling of a Swirling Fluidized Bed

Mohammad Kamil, Mokhtar Yusoff, Vijay Raghavan*

*Department of Plant and Automotive
Faculty of Mechanical and Manufacturing Engineering
Kolej Universiti Teknologi Tun Hussein Onn, 86400 Parit Raja, Johor, Malaysia
Tel: +60-7-4537794, Fax: +60-7-4536080, E-mail: vijay@kuittho.edu.my*

Abstract

Among the many variants of fluidized beds, the swirling fluidized bed is a recent entrant and offers certain advantages over conventional fluidization, such as low distributor pressure drop, higher quality of fluidization, freedom from gas bypassing as bubbles and the ability to handle odd-shaped particles. The paper presents an analytical formulation of a swirling fluidized bed and its solution by discretization of the governing equations in order to obtain its hydrodynamic characteristics. The analytical model uses a two-dimensional formulation, which enables prediction of the radial variation and axial variation of the hydrodynamic characteristics. The model developed is based on the principles of angular momentum conservation and moment equilibrium. Hydrodynamic characteristics that have been considered are the angular velocity of particles, velocity of the gas, angle of gas injection and the bed pressure drop. The results predicted with the two-dimensional model are then compared with the published experimental results. A good qualitative and quantitative agreement is established.

Keywords:

Fluidized Bed, Novel Variant, Swirling, Hydrodynamics, Modeling,

1. Introduction

Fluidization is a process by which solid particles are made to behave like a fluid, by being suspended in a gas or liquid [1]. The classical application of fluidized bed was in fluidized catalytic cracking in petroleum refining, which till today is the preferred technique of cracking the heavier hydrocarbons. Fluidization has applications in many other industrial processes including heat recovery, heat exchangers, gasification, combustion of solid fuels and particle processing such as drying, cooling and heating [2]. There have been many attempts to improve fluidized bed performance and to have different variants of its operation such as tapered fluidized bed, centrifugal fluidized bed, and vibro-fluidized bed.

One of the recent developments in providing the variant in fluidized bed operation is the swirling fluidized bed [3]. In swirling fluidization, in contrast with conventional fluidization, the fluidizing medium enters the bed at an inclination to the horizontal. It is directed thus by a suitable design of distributor. The swirling motion imparts special characteristics to the bed, which are considerably different from that for conventional fluidized beds.

As the particles move in a swirling motion, the centrifugal force causes the particle to be thrown towards the column wall, which creates a dead-zone at the center of the

bed. In industrial applications, a cone is placed at the center of the bed to prevent energy losses due to the dead-zone. Different shapes of center-body will have different effects on the bed characteristics. Though the application of the system has been widespread, the study of the process is scanty. In the present work, an analytical model with a 2-dimensional formulation is used to predict the influence of a cone at the center of the bed on bed behavior and hydrodynamics in the swirling regime.

2. Literature Review

The swirling fluidized bed is a relatively new invention. Among the early attempts in predicting the hydrodynamic characteristics of a swirl bed is the work of Wellwood [3]. The idea was to predict the slip velocity using an analogy to thermodynamics and also to identify the key parameters and their relationship with slip velocity. The thermodynamic analogy used is based on the observed similarities between microscopic molecules and macroscopic particles. It was concluded that the thermodynamic analogy of dilute gas-solid systems is a useful means of describing the slip velocity of toroidal beds. In view of the coincidence of similarity in velocity relationships, it is appropriate to further examine existing hydrodynamic models and make necessary modifications for toroidal bed applications. Shu et al [4] then developed a systematic assessment of the hydrodynamic behavior of the toroidal fluidized bed and a mathematical model to estimate particle retention time in the bed.

The study of swirling fluidized beds was taken further as

* Corresponding Author. E-mail: vijay@kuittho.edu.my,
Tel: +60-7-4537794, Fax: +60-7-4536080

Sreenivasan and Raghavan [5] and Vikram et al [6] developed an analytical model on the hydrodynamics of swirling fluidized beds. With the assumption that the angular velocity of particles remains constant across the bed height, the model gives reasonable prediction of bed pressure drop. Being a simple model, the model cannot predict the variation of the angular velocity and pressure drop with bed height and radial distance.

3. Analytical Model

3.1. Description of the System

The swirling fluidized bed system considered in developing the present model consists of two major sections: a) distributor plate, and b) bed column. Distributor blades are arranged in an annular pattern on the distributor plate. All the blades are set at the same specified angle to the horizontal. The gap between two adjacent blades and normal to their plane forms the flow area for the air supplied to the system. The velocity of the jet entering through the blade opening has two components; v_r , the tangential component responsible for the swirl motion and v_z , the vertical component responsible for the fluidization. The gap between two blades is proportional to the radius, thereby creating a trapezoidal opening for air flow.

A disadvantage of having a full-width spiral distributor is that there would be a dead zone at the centre of the column, as the bed particles are thrown towards the wall by the centrifugal force. To overcome this disadvantage, the bed column is designed to have a centre body which is located at the centre of the column / distributor. Figure 1 shows the construction of the bed.

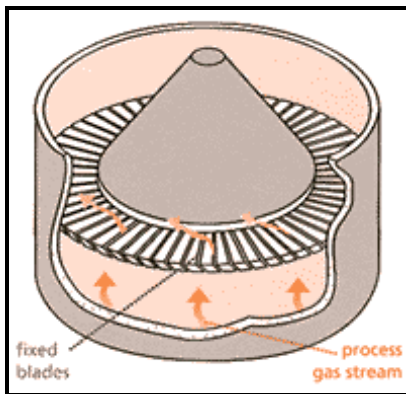


Figure 1: Construction of the swirling fluidized bed

3.2. Assumption

The major assumption involved in the formulation of the model is that the bed operates as an axi-symmetric system, which implies that the bed characteristics are a function of radial and axial distance. To determine the bed characteristics, the bed is discretized as a stack of discs in the axial direction. The discs are further considered to be composed of a series of rings at different radii. Figure 2 shows graphically, how the bed was treated. Elemental dimensions, Δh of the discs and Δr of the rings, are considered to be equal to the particle diameter, d_p .

Further assumptions applied in the formulation are listed below:

- All particles are spheres of uniform size
- All particles in a given ring have the same velocity
- The bed is in swirling regime
- The gas velocity at the distributor exit is uniform

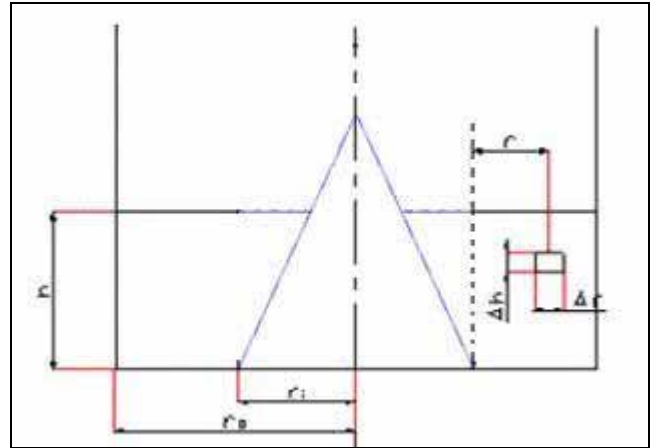


Figure 2: Schematic of swirling bed

3.3. Principle

Considering a ring as an element, the equation of force equilibrium, momentum equation and equation of continuity are applied to every ring in the bed. As stated earlier in the previous section 3.1, the flow entering the system consists of two components, the tangential and vertical components. Assuming the bed is in steady state swirling condition, the equation of force equilibrium for the tangential component can be written semantically as

$$\text{Drag force on the ring of particles due to gas flow} = \text{Opposing frictional force}$$

The drag force on the ring can be estimated as

$$F_{\text{Drag}} = n_b * \left(\frac{1}{2} C_D \rho_g \left(\pi \frac{d_p^2}{4} \right) \left[V_g \cos(\theta) - V_p \right]^2 \right)$$

The term n_b on the right hand side of the equation is to account for the drag force exerted on the ring, by the gas entering the ring from the total number of openings, n_b . Hence the drag force exerted on the ring by the total gas flow is n_b times greater than the drag force exerted by gas entering from one opening.

Since the model assumes all the particles to be spheres of uniform size, the value of drag coefficient, C_D can be taken as 0.6 from Stoke's solution with Oseen's correction for a sphere. As the ring of particles rotates in the same direction as the gas, velocity of gas applied in the drag force formula is taken as a relative velocity between the gas and the ring.

As the ring of particles swirls, two components of forces orthogonal to each other are generated, which constitute forces opposing the swirling. The first component is the downward force due to the bed weight which will cause a friction force at the bottom surface of the rings. The second

one will be the centrifugal force, acting normal to the column wall which will contribute to a frictional force between the wall and the outermost ring.

The opposing frictional force can be written semantically as

$$\begin{aligned} \text{Opposing} &= \text{Weight frictional force} + \\ \text{frictional force} &\quad \text{Centrifugal frictional force} \end{aligned}$$

The frictional force acting on the ring and directed opposite to the swirl can then be presented as

$$F_f = \mu m_r \frac{V_p^2}{r}$$

The friction coefficient, μ is considered as a tuneable value in the analysis. This value can be modified as we attempt to obtain a good agreement between the analysis and experimental results. Assuming all the particles to be contacting each other, the mass of the ring can be determined as

$$m_r = 2\pi r \frac{\rho_p \pi d_p^2}{4}$$

The equation of force equilibrium of the system can be given as

$$\left(\frac{8\mu m_r}{n_b C_D \rho_g \pi d_p^2} - 1 \right) V_p^2 + (2 V_g \cos(\theta)) V_p - V_g^2 \cos^2(\theta) = 0$$

From the above equation, the particles velocity in a given ring can be determined as a function of the inlet gas velocity, V_g and the angle, θ at which the gas penetrates into the control volume. For the rings within the first disc in the system, the inlet gas velocity and the entering angle can be determined from the distributor outlet condition. As the model assumes that the gas velocity at the distributor exit is uniform, it can be formulated by applying the continuity equation at the inlet and outlet of the distributor. The gas velocity at the distributor exit can be given by

$$V_j = \frac{U_s A_p}{n_b A_b}$$

Knowing the inlet gas velocity and the angle of entry of gas, equation of force equilibrium can be solved for the rings consisting of the first disc. For the rings located above the first disc, further information on the gas velocity and entering angle must be obtained to solve the equation. Thereby, it is necessary to determine the value of gas velocity and angle at the exit of each ring to get the inlet condition for the next upper ring. The expression for V_g and θ can be determined from the equation of continuity and momentum equation.

Considering a control volume of a ring and incompressible flow of gas, the gas velocity at the exit of the ring can be determine by continuity equation for the vertical component of gas flow, which can be written as

$$(V_g + \Delta V_g) \sin(\theta + \Delta\theta) = V_g \sin(\theta)$$

The momentum equation of the ring can be employ to determine the exit angle of gas velocity. It can be written as

$$\cot(\Delta\theta) = \frac{\Delta m V_g \sin(\theta)}{F_f} \left[\cot^2(\theta) - \frac{F_f \cot(\theta)}{\Delta m V_g \sin(\theta)} + 1 \right]$$

where Δm , is the mass flow rate entering the ring and can be expressed as

$$\Delta m = n_b \rho_g V_j \left[a + \frac{(b-a)}{(r_o - r_i)} (r - r_i) \right] \Delta r$$

where a and b are the blade opening gaps at the inner and outer radii respectively while r_o and r_i are the inner and outer radii of the blade opening.

As a further extension of the analysis, the bed pressure drop is evaluated. Previous models [5,6] proposed that the bed pressure drop in a swirling fluidized bed increases as the angular velocity of the bed increases. It is proposed that, in addition to the weight of the bed, the centrifugal force of the bed acting normal to the column wall also contributes to a downward acting force during the bed expansion [6].

Applying this principle to the system, the bed pressure drop can be written as

$$\Delta P^* (2\pi r d_p) = m_r \left(g + \mu \frac{V_p^2}{r} \right)$$

where the first term on the left hand side represents the weight of the bed, and the second term represents the downward force contributed by the centrifugal force while the bed is expanding.

3.3. Solution

Parameters such as exit gas velocity, exit angle of gas and mass flow rate of a given ring need to be determined to solve the model explicitly where the outlet condition of the previous disc will determine the inlet condition of the next disc.

The finite change on the mean flow area from one disc to the next disc corresponding to the bed geometry will account for setting the inlet parameters for the upper rings. Considering, i as the number of rings in the previous layer, the inlet velocity for the vertical direction can be written as

$$(V_g + \Delta V_g) \sin(\theta + \Delta\theta) = \frac{1}{\sum_{i=1}^i A} \left[V_g \sin(\theta) \right]$$

From the above equation, the inlet velocity of the next layer is proportional to the outlet gas velocity by a factor which is termed as vertical stretching factor sf_v and the value can directly obtained from the flow area ratio between the succeeding and preceding discs.

For the horizontal component of the gas flow, the inlet gas velocity can be expressed as

$$(V_g + \Delta V_g) \cos(\theta + \Delta\theta) \Big|_{i+1} = \sum_1^i V_g \cos(\theta) - sfh \sum_1^i (V_g + \Delta V_g) \cos(\theta + \Delta\theta) \quad (v)$$

where sfh is horizontal stretching factor. sfh is a tuneable parameter which is determined to give a smooth regression of velocity profile along the radial distance. This is an important aspect in the solution procedure to avoid erroneous results and accumulation of errors. Solving both the vertical and horizontal component, the inlet gas velocity and entering angle of gas velocity for the new layer can be determined from ring to ring.

4. Results and Discussion

4.1. Quantitative Agreement

The proposed model is capable in determining a number of hydrodynamic characteristics in swirling fluidized bed operation including gas velocity, the inclination of gas velocity, the particle linear velocities, the particle angular velocities and the bed pressure drop. Among all of the bed characteristics, bed pressure drop was the only parameter that can easily be determined by experimental approach. For the validation of the model, the analytical bed pressure drop result determined by the model is compared with the experimental result of the bed pressure drop given in the previous work by Binod and Raghavan [5]. Figure 3 shows the comparison between the analytical result and the experimental result on the bed pressure drop.

As can be seen from Figure 3, good agreement between the analytical result and the experimental result has been achieved. It shows that the proposed model is valid. A small error ($\pm 10\%$) between the experimental and analytical results shows that further improvements have to be made on the model.

4.2. Qualitative Agreement

In general, the results confirm the visual observations on a swirling fluidized bed. The main visual observations are:

- (i) The bed swirls uniformly, i.e., at a fairly constant angular velocity from the inner radius to the outer radius at the top surface of the bed, which is the only observable surface.
- (ii) The swirling velocity is a maximum at the distributor, decaying progressively towards the upper surface of the bed. This decline is due to the fact that the angular momentum of the gas is continuously transferred to the bed particles.
- (iii) The angular velocity of the particles is a maximum at the distributor. In other words, the particle swirling is at its most vigorous at the distributor and reduces with bed height.
- (iv) The steady swirling of the bed suggests that the energy given by the gas causes the bed particles to accelerate from an initial velocity of zero. Once steady operation has been

reached, the energy of gas transferred to the particles is steadily consumed against friction, viz., friction at the wall and gas-particle friction.

(v) The gas enters the bed at a low angle and gradually the inclination of its velocity will increase. At a certain bed height, it operates as a two-layer bed, the lower part swirling and the upper part of the bed operating as a bubbling fluidized bed.

(vi) The bubbling (non-swirling) mode appears to coincide with the disappearance of all angular momentum of the gas as it flows through the bed, dissipating its energy against bed friction.

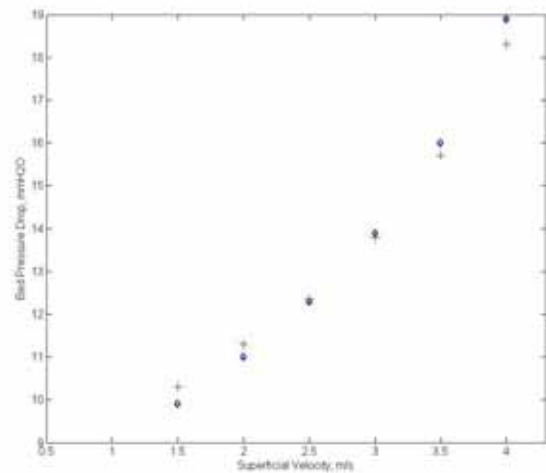


Figure 3: Comparison between the analytical result and the experimental of the bed pressure drop

In the present two-dimensional analysis of the swirling bed, the results also confirm all the observations stated above. There is a variation in angular velocity and linear velocity of gas, and peripheral velocity of particles in the radial direction as well as the axial direction. This justifies the two-dimensional approach to the problem. The qualitative agreement of the obtained results against visual observations vindicates the consideration of 2-D variations in the present work. The physical plausibility of the results confirms that the model and the results are realistic. These variations cannot be verified by measurements except at the upper surface of the bed. Such measurements are quite complex, requiring tracking of particles and are expected to be taken up at a later part of our studies.

The characteristics of the swirling fluidized bed in the current study are estimated for the following parameters: inlet blade angle of 12 degrees from the horizontal, bed weight of 0.5 kg, particle diameter 3.5mm, superficial gas velocity of 4 m/s, bed outer diameter of 310mm, conical base diameter of 200mm and the converging angle of the cone is 37.6 degrees to the vertical.

Figure 4 presents the variation of particle velocity in the axial direction. It can clearly be seen that the velocity of the particles is a maximum at the distributor and decays as we progress in the axial direction, confirming the visual observations (ii,iii). Figure 5 shows the variation of the

inclination of gas velocity with the horizontal. This happens due to faster decay of the angular momentum at the inner radius due to lower gas flow rate. On account of its lower angular momentum, the gas flow at the inner radius suffers a greater change in gas flow direction towards the vertical.

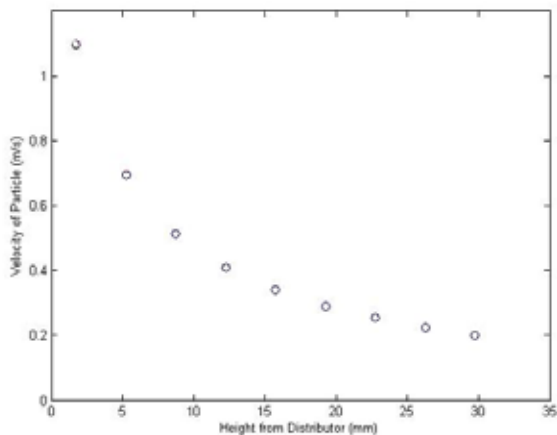


Figure 4: Variation in Velocity of Particles in Axial Direction

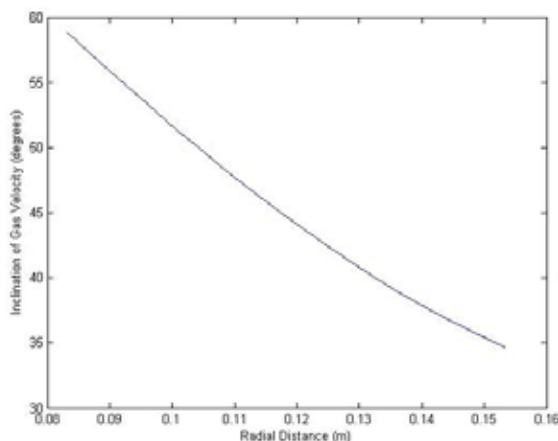


Figure 5: Variation in Inclination of Gas Velocity in Radial Direction

Figure 6 shows the inclination of gas velocity to the horizontal. Owing to the fact that the horizontal component of the gas velocity continuously decreases by momentum transfer to swirling particles, the inclination of gas with the horizontal goes on increasing as it travels up the bed. The axial variation shows that the angular momentum of gas is lost rapidly on entering the bed and tapers off as we move upwards in the bed. The inclination of gas velocity will increase as we move upwards through the bed and will tend to reach 90 degrees at a certain bed height. At this point, the gas has lost all its angular momentum and will operate as a two-layer bed as described in visual observations (v,vi).

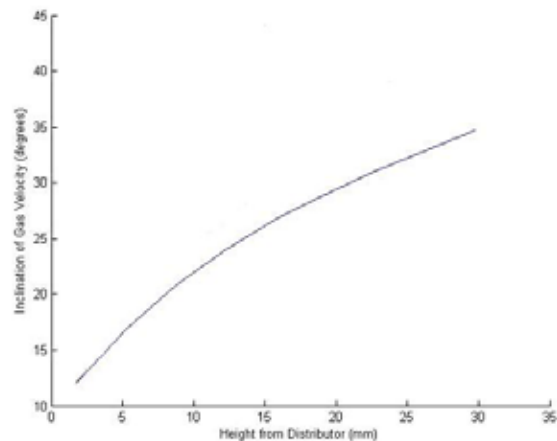


Figure 6 : Variation in Inclination of Gas Velocity in Axial Direction

5. Conclusion

A two-dimensional model is proposed in the study. The proposed model is capable to determine the hydrodynamic characteristics of a swirling fluidized bed. The analysis is performed based on the conservation of angular momentum. Good agreements have been achieved between the analytical result and the published experimental result with error of $\pm 10\%$.

The two-dimensional model attempts a more realistic prediction of the hydrodynamic characteristics of swirling beds. The validation of the two-dimensional formulation is only carried out in terms of bed pressure drop, as it is well nigh impossible to determine other hydrodynamic characteristics experimentally.

References

- [1] J.R. Howard (1989), 'Fluidized Bed Technology: Principles and Applications', Adam Hilger, New York.
- [2] C.K. Gupta and D. Sathiyamoorthy (1999), 'Fluid Bed Technology in Materials Processing', CRC Press, New York.
- [3] G.A. Wellwood (1997), 'Predicting the Slip Velocity in a TORBED Bed Reactor Using an Analogy to Thermodynamics', 14th Int. Conf. Fluid. Bed Combust., Vancouver, Canada, 618-628.
- [4] J. Shu, V.I. Lakshmanan and C.E. Dodson (2000), 'Hydrodynamic Study of a Fluidized Bed Reactor', Chemical Engineering and Processing, 39, 499-506.
- [5] B.Sreenivasan and V.R. Raghavan (2002), 'Hydrodynamics of a Swirling Fluidised Bed', Chemical Engineering and Processing, 41, 99-106.
- [6] G. Vikram, V.R. Raghavan and H. Martin (2003), 'A model for the Hydrodynamics of Swirling Fluidised Beds', 4th International Conference for Conveying and Handling Particulate Solids, Budapest, Volume 1, 7.1

Analysis of Flow and Heat Transfer over Louvered Fins in Compact Geometries

Yong Kok Wee¹, Vijay R. Raghavan^{1*}

¹ Kolej Universiti Teknologi Tun Hussein Onn,
86400 Parit Raja, Batu Pahat, Johor, Malaysia.
Tel: +60-126793735, E-mail: yong_kok_wee@hotmail.com

Abstract

The aim of the present study is to determine the louver inclination that has the greatest influence on fin heat transfer. Developing an analytical model, the flow and heat transfer from compact louvered fins could be analyzed. To standardize the mesh size and also the solution approach, validation is obtained. By fixing all the hydrodynamic variables except for louver angle, the geometries are investigated using computational methods based on finite volume analysis. Air velocities in the practical range of 3 – 10.5 m/s are considered. The large amount of data obtained for the range of louver inclination from 15 degrees to 35 degrees, are analyzed to obtain corroboration of the methods followed. The processed data can be plotted as dimensional parameters, in the form of (i) Nusselt number, (ii) friction factor (iii) the ratio of heat transfer to pumping power as function of Reynolds number. As anticipated from the earlier experiments, it is concluded that, among all the fin and louver parameters, the inclination of the louver could have the greatest influence.

Keywords:

Compact Heat Exchanger, Louver Inclination, Numerical

1. Introduction

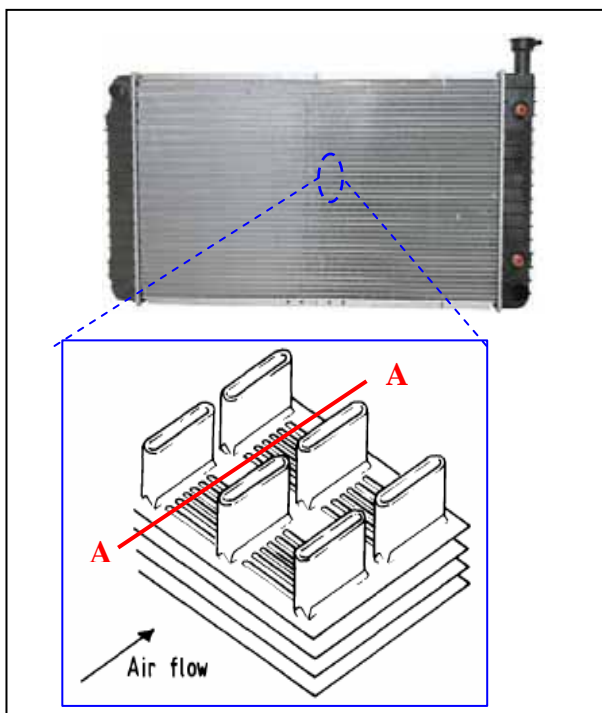


Figure 1: Louvered fins in the radiator.

For new designs of a compact heat exchanger for rejection of heat in automotive applications such as radiators, it is necessary to predict the heat transfer capability and the associated flow resistance of the system. When a large heat transfer surface per unit volume is desired, the heat exchanger will be of a compact design incorporating a close-pitched stack of fins.

The preferred fin geometry in radiators is the louvered fin, because (i) its manufacture can be automated (ii) it can be accommodated to give great compactness in radiator, as compactness is important in an automobile, and (iii) unlike other types, it gives considerable heat transfer enhancement without a correspondingly large pressure drop.

A study of this kind of finning is useful in compact heat exchangers that have been recognized in aerospace, automobile, and other industries. Of the many types of fins that have been studied in compact geometries, such as strip fin, offset fin, wavy fin etc., the louvered fin appears to be the most suitable for automotive applications.

The hydrodynamic characteristics of a fin stack depend on a number of variables: fin pitch, fin thickness, number of louvers, louver pitch, louver height and louver angle. In the process of designing compact heat exchangers, an analytical model is a crucial development to enable prediction of heat transfer and pressure drop.

*author for correspondence: vijay@kuittho.edu.my

2. Past Studies of Relevance

Davenport [1] suggested that the fluid flow tended to be largely down the gap between fins at very low Reynolds numbers. This gradually changed to almost complete alignment with the louvers as Reynolds number was increased. He conjectured that at low Reynolds numbers the laminar boundary layers on the louvers are sufficiently thick to effectively block off the gaps between adjacent louvers in the same fin. Confirmation of this idea comes from Achaicia and Cowell [2] where they determined how the degree of flow alignment changed upon Reynolds number and the geometric parameters. They used a two dimensional finite-difference analysis of the fully developed periodic flow situation in an infinite louver array with infinitely thin louvers. A mean flow angle was determined by the cotangent of the flows into and out of the west and north faces, respectively, of the fully developed periodic flow domain. They proposed equation (1) to describe the behavior.

$$\beta = 0.936 - (243/\text{Re}_L) - 1.76(F/L_p) + 0.995\alpha \quad (1)$$

Zhang and Tafti [3] found that the thermal wake effects could be expressed as a function of the flow efficiency (measurement of how duct- or louver-directed the flow is) and the fin to louver pitch ratio. Moreover, they state that neglecting thermal wake effects at low flow efficiencies can introduce errors as high as 100% in the heat transfer. The advantage of computational analysis in simulating thermal wake effect could provide predictions that are better than large scale experiment.

In the CFD analysis of Perrotin and Clodic [4] to determine heat transfer and pressure drop characteristics and local information analysis, they simulated 2D models, with constant wall temperature boundary condition; the error of the heat transfer co efficient is up to 80% compared to the experimental data. However, 3D models take into account the fin thermal conductivity and the tube geometry and the heat transfer co-efficient calculated is much closer to the experimental data (13%). Thus, 3D CFD model simulation would give a closer and more accurate prediction.

3. The Approach and Methods

3.1 Procedure of the Study

In the description of the system in *figure 2*, the crucial needs of doing such study are understood and the scope is defined within the limitations of the software and hardware. In accordance with the geometry and its functions, a proper control volume is laid out. In the analysis of the research problem, the equations of mass conservation, momentum conservation and energy

conservation are understood and the conditions of the study are formulated.

From the previous published study by Achaicia and Cowell [2] on the determination on how the degree of flow alignment depended upon Reynolds number and the geometric parameters, the model in this study is validated with the equation (1). This is done in order to make comparison of the model with established results and to standardize the choice of mesh size and its distribution and the solution procedure.

In finding the louver angle that has the greatest influence on heat transfer in the fin stack, the geometry boundaries are redefined for the problem. Repeating the solution procedure with suitable boundary conditions and by varying louver angle and air velocity, a large amount of data are obtained for the range of louver inclinations from 15 degrees to 35 degrees.

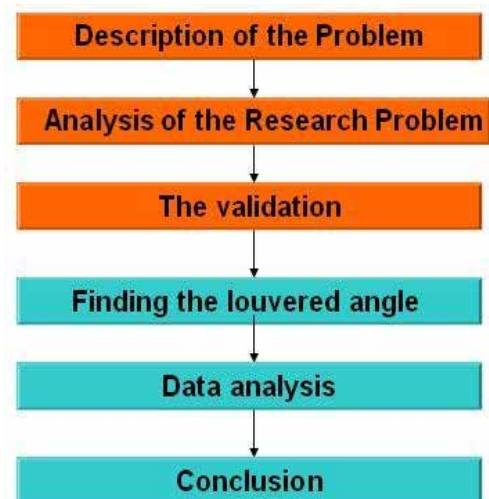


Figure 2: The procedure of the study

In the final analysis of data, graphs are to be plotted as dimensional parameters, in the form of (i) Nusselt number, (ii) friction factor (iii) the ratio of heat transfer to pumping power as function of Reynolds number. Discussion and conclusions are presented at the end of the study.

3.2 Description of the Control Volume



figure 3: Cross section of A-A

The cross section of A-A in the figure 1 is shown in figure 3. The control volume boundary of the model in CFD covers only the left half of the full stack of fins as shown in *Figure 4*.

The flow is assumed to be laminar which is reasonable at the considered Reynolds number Re_{Lp} ranging from 300

to 1200. Experimental observations of Antoniou et al. [5] have shown that the flow is laminar for Reynolds numbers up to 1200.

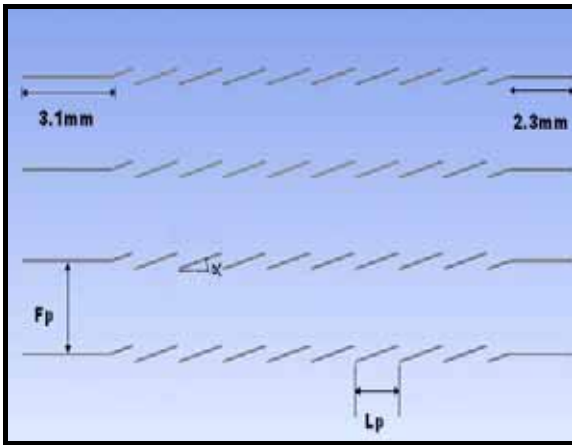


Figure 4: Geometries of the control volume

The Reynolds number is based on louver pitch as shown in equation (2).

$$Re_L = \frac{VL_p}{\nu} \quad (2)$$

3.3 The Governing Equations

The set of equation is solved by the Ansys CFD are the unsteady Navier-Stokes equations in their conservation form. The simulations are performed as incompressible flow. The continuity equation is simplified as Eq. (3):

$$\nabla \cdot U = 0 \quad (3)$$

The momentum conservation equation is Eq. (4). Simulations in the ideal gas domain are performed as fully developed periodic flows. Relatively, the model is too small for taking gravitational acceleration into consideration.

$$\rho \nabla \cdot (U \otimes U) = \nabla \cdot (-p\delta + \mu((\nabla U) + \nabla U)^T) + S_M \quad (4)$$

There is no energy generation and work done by the fluids. The energy conservation equation is Eq. (5).

$$\rho \nabla \cdot (U h_{tot}) = \nabla \cdot (k \nabla T) + S_E \quad (5)$$

h_{tot} is defined as the total specified enthalpy and is given in the term of specific static enthalpy, h , by:

$$h_{tot} = h + \frac{1}{2} U^2 \quad (6)$$

where h is a function of pressure.

$$h = h(p, T) \quad (7)$$

3.4 The Boundary Condition

The air flows into the west surface and the air also entrained into the control volume from the south opening in figure 3. The air then flows out through the north and east opening. The inlet velocity is specified and ranged from 3 m/s to 10.5m/s.

$$U_{Inlet} = \begin{bmatrix} U_x \\ U_y \\ U_z \end{bmatrix} = \begin{bmatrix} U_x \\ 0 \\ 0 \end{bmatrix} \quad (8)$$

The opening boundary condition is specified with a relative pressure of zero. The pressure value is taken to be static pressure for both the inflow and outflow. This option is useful for situations in which the main flow tends to entrain the fluid through the boundary.

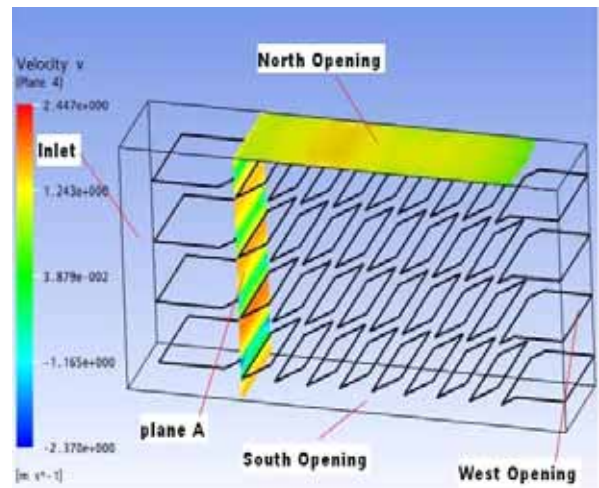


Figure 5: Boundary conditions and regions

The body of the fins is set as non-slip wall where the velocity at the wall is zero.

$$U_{wall} = 0 \quad (9)$$

The heat flux at the non-slip wall is defined as

$$q'' = h_c (T_{fluid} - T_{nearwall}) \quad (10)$$

The T_{fluid} is the specified boundary temperature with the co-efficient, h_c , and $T_{nearwall}$ is the temperature at the internal near-wall boundary element centre node.

In finding the louvered angle that has the greatest influence on heat transfer, one side of the fins is defined as adiabatic wall where no heat transfer is allowed across the defined boundary.

$$q_w = 0 \quad (11)$$

The adiabatic wall condition is defined on the same coordinate plane as symmetrical plane. The velocity component normal to the symmetrical plane is set to be zero.

$$U_n = 0 \quad (12)$$

Also, the scalar variable gradients normal to the symmetrical plane are also set to zero.

4. The Results and the Progress

According to the definition of mean flow angle by Achaicia and Cowell [2], only the louvered surfaces are considered in the validation. Thus, the velocity is taken starting from plane A as shown in figure 5.

Table 1

Louver angle	Velocity Plane A (ms ⁻¹)	CFD Mean flow angle	Error
15°	10.32600	13.84934249	6.59%
15°	8.833680	13.89116948	7.23%
15°	7.346060	13.93206800	8.00%
15°	5.865670	13.95726801	8.88%
15°	4.391410	13.98057198	10.22%
15°	2.924950	13.59994676	9.54%
20°	10.101400	18.83909404	4.88%
20°	8.639200	18.47318083	3.07%
20°	7.183600	18.41446423	3.06%
20°	5.734710	18.26429277	2.70%
20°	4.283900	18.10635160	2.61%
20°	2.851210	17.44046575	0.41%
25°	9.995210	23.21112534	1.20%
25°	8.555530	23.11964260	0.98%
25°	7.112240	23.05256203	0.93%
25°	5.679950	22.86065711	0.46%
25°	4.250150	22.46052354	-0.69%
25°	2.831170	21.68847921	-2.91%
30°	9.893860	27.46769748	-1.58%
30°	8.467850	27.33336565	-1.92%
30°	7.046760	27.13401593	-2.44%
30°	5.631220	26.79859026	-3.35%
30°	4.224700	26.17662851	-5.12%
30°	2.814070	25.34699865	-7.18%
35°	9.783690	31.29407032	-4.82%
35°	8.378160	31.12030660	-5.24%
35°	6.975880	30.87363815	-5.82%
35°	5.577430	30.41857669	-6.97%
35°	4.182460	29.69159527	-8.80%
35°	2.788970	28.69847510	-11.08%
Absolute mean error			4.62%

The validation results agreed with equation (1) within an absolute mean of error of 4.62% and a maximum deviation of 11.08% as shown in table 1. Errors tended to

be larger at low Reynolds number except for louver angle of 20°. In the range of louver angles of 20° to 30°, the agreement is excellent, with an absolute mean error of 2.56%. The study to determine the extent of influence of the inclination of the louver angle in comparison with other variables is in progress.

From the validation results, it is shown that, whether 2 dimensional finite elements or finite volume method is used, the difference is small for the problem solved. For minimizing the errors in any typical simulations of flow and heat transfer over the louver fins, the Reynolds number of the flow should be higher than 300 (or 3ms⁻¹ for an ideal gas for the range of practical louver pitches).

Acknowledgments

The authors would like to acknowledge with appreciation, the help and support from the Faculty of Mechanical Engineering, Kolej Universiti Teknologi Tun Hussein Onn, in providing all facilities for this research.

Nomenclature

- dot product.
- ✓ louver angle.
- β mean flow angle.
- ☛ distance traveled by fluid.
- F fin pitch.
- k thermal conductivity.
- L louver pitch.
- h enthalpy.
- h_c heat transfer co-efficient.
- h_{tot} total specified enthalpy.
- ρ density of the fluid.
- P pressure.
- q_w heat transfer rate into wall.
- q heat flux.
- Re_L Reynolds number based on Louver pitch.
- S_M defined as the momentum source.
- S_E defined as energy source in a control volume.
- T temperature.
- U velocity field in vector form.
- U_{inlet} velocity field in vector form at the inlet.
- U_n is velocity normal to the plane.
- U_{wall} velocity field in vector form near the wall.
- U_x velocity at x-direction.
- U_y velocity at y-direction.
- U_z velocity at z-direction.
- μ dynamic viscosity.
- V velocity inlet.
- ν kinematics viscosity.
- ∇ 'del' and defined as $[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}]$.
- ⊗ cross product.
- $()^T$ matrix transposition.

Reference

- [1] Davenport, C. J., Heat Transfer and Fluid Flow in Louvered Triangular Ducts, Ph.D. Thesis, CNAAB, Lanchester Polytechnic, Coventry, UK, 1980.
- [2] Achaichia, A, and Cowell, T.A., Heat Transfer and Pressure Drop Characteristics of Flat Tube and Louvered Plate Fin Surfaces, *Exp. Thermal Fluid Sci.* 1, 147 -157, 1988.
- [3] Zhang, X., Tafti, D.K, Classification and effects of thermal wakes on heat transfer in multi-louvered tubes, *International Journal of Heat and Mass Transfer* 44 (2001) 2461–2473.
- [4] Perrotin, T., and Clodic, D., “ Thermal-hydraulic CFD study in louvered fin-and-flat-tube heat exchangers”, *International Journal of Refrigeration* 27, 2004, Elsevier Ltd, 422–432.
- [5] Antoniou, A. A., Heikal, M.R., and Cowell, T. A., Measurements of local velocity and turbulence levels in array of louvered plate fins. Proceedings of the 9th International Heat Transfer Conference on Numerical Methods in Laminar and Turbulent Flow, Montreal, 1987, p. 482–495.

Experimental Study on Manoeuvring Characteristics of a Planing Hull

A. Haris Muhammad^{1*}, Adi Maimun¹, Omar Yaacob¹, Agoes Priyanto¹

¹ Faculty of Mechanical Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5535706, E-mail: andi_haris@ft.unhas.ac.id

Abstract

A patrol vessel usually operate in coastal and interior waters, in good to harsh visibility condition and frequently involved in speedy and close proximity operation requiring a deep knowledge of the manoeuvring characteristic of this vessel. This paper describes an experimental investigation into the manoeuvring characteristics of the patrol vessel. The vessel was fitted with 3 spray-strakes along her bare hull in port and starboard side. The experimental studies measured the hydrodynamic derivatives using a planar motion mechanism (PMM). By using the time simulation approach included the hydrodynamic derivatives, the manoeuvring characteristics (turning trajectory and zig-zag manoeuvre) were calculated. The results shown that the turning trajectory of vessel with 3 spray-strakes has larger tactical diameter than the bare hull one.

Keywords:

Manoeuvring, Planing Hull, Spray-Strake, Hydrodynamic Derivative, Planar Motion Mechanism (PMM)

1. Introduction

Since the application of planing hulls in role as patrol, coastguard, survey or naval vessels have increased considerably over the last two decades, the improvement of behavior of these vessel have been an interest research topic, especially to find the manoeuvring qualities in sea operation.

Ship manoeuvrability is an important requisite to be satisfied IMO limit criteria when designing a new ship. The problem of foreseeing a ship's behaviors at sea is yet unresolved, especially for high speed craft (HSC), which no reliable procedure exists so far. However, on assessing their manoeuvring qualities, HSC can be used to satisfy the same IMO manoeuvring limit criteria in calm water as conventional ship [1].

As a matter of fact, the increasing high speeds in navigation that can now be attained (well over the threshold value of forty knots, even with sizes of up to one hundred meters in length) make their control at sea particularly difficult. Coccoli [2] and Soares [3] have investigated on the manoeuvring by full scale sea trial on an HSC, and found that the characteristic of tactical diameters (D_T) and advances (A_D) manoeuvre value of the craft obtained was over the IMO limit criteria.

Many efforts for the improvement on the ship manoeuvring have been done through the selection of parameters that influence their characteristics, such as additional mass because of the forward speed. Other parameters also influence their manoeuvring performances such as deadrise angle, trim angle and draught variations [4] [5]. Besides, that phenomenon can be also influenced by the effect of some appendages (bilge and central keel) that are attached on the bottom [6].

Starting from the above considerations, this research concentrates on the use of 3 spray-strakes appendages which attached along the bare hull of the 22 m planning vessel. The experimental studies measured the hydrodynamic derivatives of the planing vessel without/with 3 spray-strikes using a Planar Motion Mechanism (PMM). The manoeuvring characteristics (turning trajectory and zigzag manoeuvre) were calculated by using the time simulation programs with the input of the hydrodynamic derivatives.

2. Mathematical Model

In manoeuvring theory [7, 8 and 9], an axis system fixed in the vessel with the origin positioned at the centre of mass allow development of a suitable mathematical model to simulate its manoeuvring. The mathematical model can be described by the Equation (1), using the coordinate system in Figure 1.

$$\begin{aligned} X &= m(\dot{u} - rv) \\ Y &= m(\dot{v} + ru) \\ N &= I_{zz}\ddot{\psi} \end{aligned} \quad (1)$$

* Corresponding Author. E-mail: andi_haris@ft.unhas.ac.id, Tel: +60-7-5535706.

Where:

- m : mass of ship
- I_{zz} : moment of inertia of ship in yaw motion

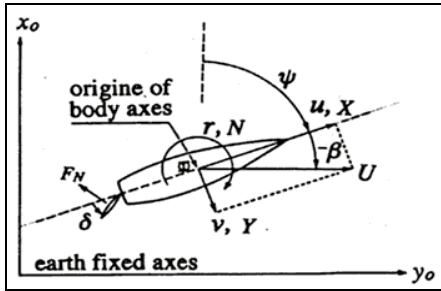


Figure 1. Co-ordinate system

The notation of u_G , v_G and r are velocity components at centre of gravity of ship (C.G), and x_G represent the location of the C.G in x-axis direction. X , Y and N represent the hydrodynamic force and moment acting on the mid ship of hull. These forces can be described separating into the following component from the viewpoint of the physical meaning.

$$\begin{aligned} X &= X_H + X_R + X_P \\ Y &= Y_H + Y_R + Y_P \\ N &= N_H + N_R + N_P \end{aligned} \tag{2}$$

Where, the subscripts H , P and R refer to hull, propeller and rudder respectively according to the concept of Mathematical Modelling Group (MMG).

2.1. Forces and Moment Acting on Hull

The forces and moments during measured in experiment can be modeled by the following equation of motion.

$$\begin{aligned} X_H &= (X_{\dot{u}} - m)\dot{u} + (X_{vr} + m)vr \\ &\quad + X_{vv}v^2 + X_{rr}r^2 \quad (\text{RESISTANCE}) \\ Y_H &= Y_0 + (Y_{\dot{v}} - m)\dot{v} + Y_{\dot{r}}\dot{r} + Y_vv + (Y_r - m)r \\ &\quad + Y_{vv}v^3 + Y_{vr}v^2r + Y_{rr}vr^2 + Y_{rrr}r^3 \\ N_H &= N_0 + (N_{\dot{v}} - mx_G)\dot{v} + (N_{\dot{r}} - I_{zz})\dot{r} + N_vv + N_r r \\ &\quad + N_{vv}v^3 + N_{vr}v^2r + N_{rr}vr^2 + N_{rrr}r^3 \end{aligned} \tag{3}$$

2.2. Forces and Moment Induced Propeller and Rudder

The forces and moments induced by propeller and rudder can be modeled by the following formula.

$$\begin{aligned} X_P &= (1 - t_p)\rho K_T D_p^4 n^2 \\ Y_P &= 0 \\ N_P &= 0 \end{aligned} \tag{4}$$

Where:

$$\begin{aligned} K_T(J_P) &= C_1 + C_2 J_P + C_3 J_P^2 \\ J_P &= U \cos \beta (1 - w_p) / (n D_p) \end{aligned}$$

Where: Where t_p : thrust reduction coefficient in straightforward moving, C_{iP} : constant, n : propeller revolution, D_p : propeller diameter, w_p : effective wake fraction coefficient at propeller location, J_p : advance coefficient, C_1, C_2, C_3 : constant.

The terms on rudder force are assumed as follows.

$$\begin{aligned} X_R &= (1 - t_r)F_N \sin \delta \\ Y_R &= (1 + a_H)F_N \cos \delta \\ N_R &= (x_R + a_H x_H)F_N \cos \delta \end{aligned} \tag{5}$$

Where, δ is rudder angle, x_R represents the location of rudder ($= L/2$), and, t_r , a_H and x_H are the interactive force coefficients among hull, propeller and rudder. K_T is the thrust coefficient of a propeller force. These are the functions of the advance constant of propeller. The dimensional of ruder forces describes as the following.

$$F_N = \frac{1}{2} \rho A_R f_\alpha U_R^2 \sin \alpha_R \tag{6}$$

Where, A_R is rudder area, f_α is the gradient of the lift coefficient of ruder, and can be approximated as the function of rudder aspect ratio Λ .

$$f_\alpha = 6.13\Lambda / (2.25 + \Lambda) \tag{7}$$

Where:

$$\begin{aligned} U_R &= \sqrt{u_R^2 + v_R^2} \\ \alpha_R &= \delta \tan^{-1} \frac{v_R}{u_R} \end{aligned} \tag{8}$$

Where:

$$\begin{aligned} u_R &= \varepsilon (1 - w) u \\ &\quad \sqrt{\mu \left\{ 1 + \kappa \left(\sqrt{1 + (8K_T / \pi J^2)} - 1 \right) \right\}^2 + (1 - \eta)} \\ v_R &= \gamma_R (v - r l_R) \end{aligned}$$

ε , κ , η and l_R in the above equations are the parameters with the rudder inflow velocity and angle. $(1 - w)$ and η are the effective propeller wake fraction and the ratio of propeller by rudder height (D_p/H)

3. Experimental Model

The experiments were conducted in the 120 meter long towing tank at Marine Technology Laboratory of Universiti Teknologi Malaysia (UTM).

3.1. The Model

The ship model was a 1/10 scale model of a 22.0 m long planing hull “Rajawali” (Figure 2). The model introduced the use of spray-stakes attached (Figure 3). The main particulars of the vessel are shown in Table 1.

Spray stakes were generated according to geometry description of the spray stake in ref. [10], which is shown in Figure 4. The spray-strike had a triangular cross-section with constant bottom with $bsr=15$ mm in model scale. The bottom angle β , between the bottom of the stake and the horizontal line is 0° , and $\zeta = 90^\circ$

Table 1. Main dimensions of planing hull

Dimension	Ship	Model
Loa (m)	22.00	2.200
Displacement, Δ (tonne)	55.16	0.551
Chine beam, B_x (m)	4.906	0.491
Projected chine length, L_p (m)	21.13	2.113
LCG forward of transom (m)	9.22	0.922
Midship bottom deadrise, β (deg)	22.95	22.95
Draft (m)	1.243	0.124
Block Coefficient, C_b	0.435	0.435
Spray-stake number	-	2 ~ 3
Speed (m/sec)	7.71	2.438

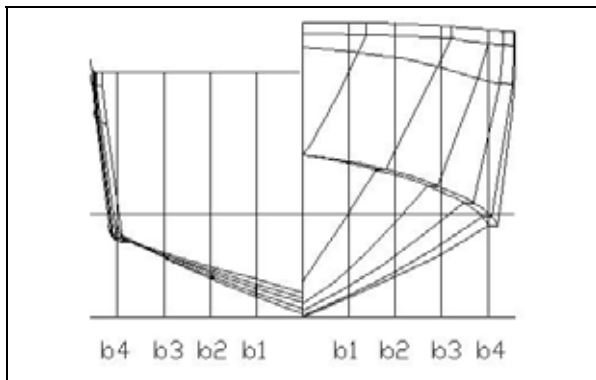


Figure 2: The original body plan

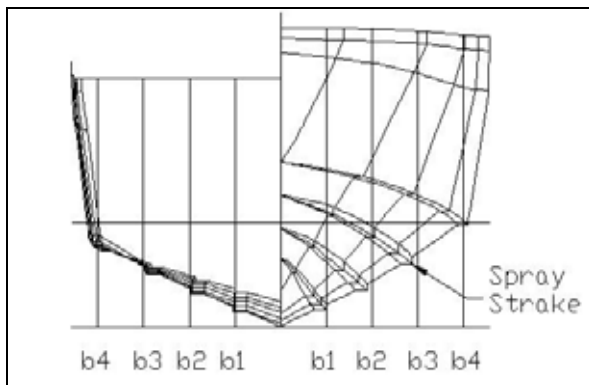


Figure 3: The body plan with spray-stake

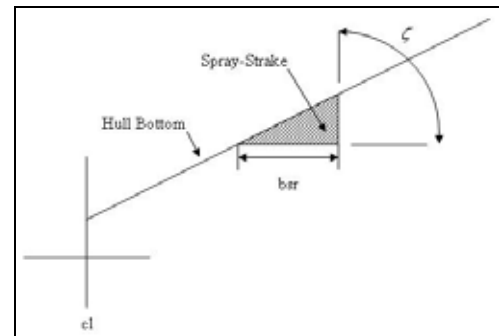


Figure 4. Type and sectional Shape of tested Spray-stake

3.2. The configuration and parameter test

The Planar Motion Mechanism (PMM) test has been widely used to obtain the hydrodynamic coefficients in vessel manoeuvring problems. Most frequently, the experiment data have been analysed by a Fourier integral method. An oscillatory motion is imposed on the model when it was towed in the tank with constant speed. The motion generated by PMM could be drift tests, pure sway, pure yaw, and yaw with drift test. The model is also free to heave and pitch, but restrained in surge, sway, and yaw motions. The input parameters of these tests are presented in Table 2.

Table 2. Input parameters of PMM test

Model Test	Drift angle (deg.)	Amp (m)	Phase (deg)
Drift Test	$0^\circ, 4.83^\circ, 9.66^\circ, 14.49^\circ$	0	0
Pure Sway	0°	0.1 ~ 0.5	0
Pure Yaw	0°	0.1 ~ 0.5	11.77°
Yaw with drift	$4.83^\circ, 9.66^\circ, 14.49^\circ$	0.05 ~ 0.20	11.77°

3.3. Experimental Result

3.3.1. The Longitudinal Forces

Figure 5 shows the results of vessel resistance. The non dimensionless force coefficient, $X'(u')$ are derived from the resistance test .

Figure 6 and Figure 7 shows the surge force measured by PMM. The results are expressed by the equation (1) with the coefficient of $X'vv$ and $X'rr$, in which the symbol \blacksquare, \bullet and so on denote the measured values for the bare hull without and with spray stakes, respectively. The full lines are approximated curve for the bare hull and the broken lines are done for hull with spray-stake.

3.3.2. Lateral Force and Moment

Figure 8 and 9, shows the lateral force and moment measured by PMM test. The results are expressed by the equation (1) with the coefficient of $Y'vvv, Y'rrr, Y'v$ and $Y'r$ (sway force coefficient) and $N'vvv, N'rrr, N'v$ and

$N'r$ (yaw moment coefficient) where the symbol \blacksquare, \bullet and so on denote the measured values for the bare hull without and with spray strakes, respectively. The full lines are approximated curve for bare hull and the broken lines are done for hull with spray-strake.

For another coefficient such as added mass coefficient in the longitudinal direction ($X'_{\dot{u}}$) was estimated from the Motora's chart [11]. Furthermore, interactive coefficient of hull-rudder-propeller prediction was calculated based on the Yoshimura Method in ref. [12]. The complete all coefficients for use in simulation have in Table 3 and Table 4.

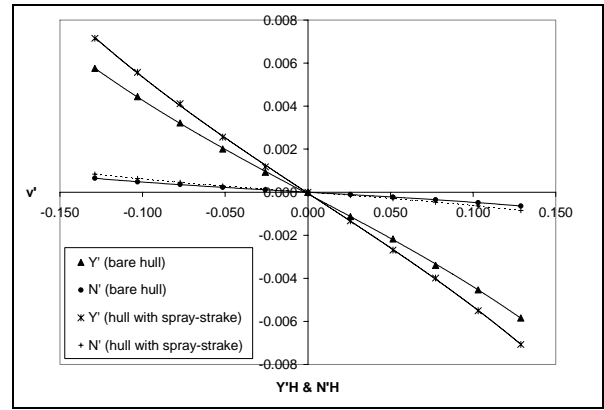


Figure 8: $Y'H$ and $N'H$ vs. v' graph

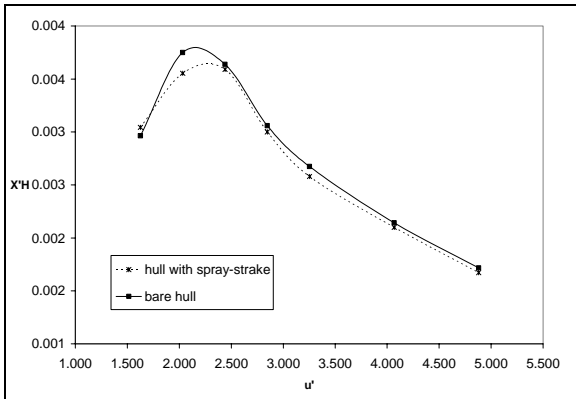


Figure 5: Non-dimensional of the resistance test

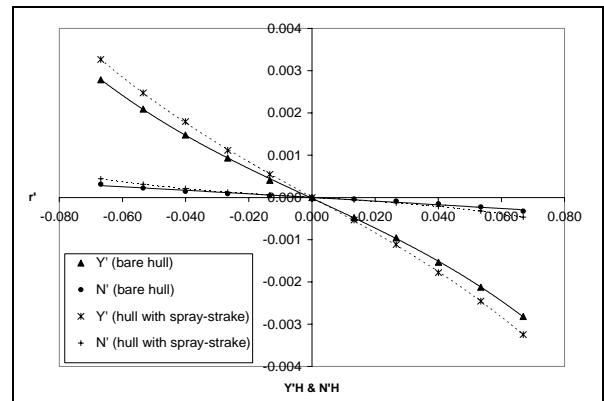


Figure 9: $Y'H$ and $N'H$ vs. v' graph

Table 3: X, Y and N Hydrodynamic derivatives

	Bare Hull	Hull with Spray-Strake
X'_{uu}	-0.0034	-0.0034
$X'_{\dot{u}} - m'$	-0.0100	-0.0127
$X'_{vr} + m'$	0.1684	0.0598
X'_{vv}	0.0224	0.0326
X'_{rr}	0.1900	0.1824
$Y'_{\dot{v}} - m'$	-0.0140	-0.0141
$Y'_{\dot{r}}$	-0.2516	-0.2962
$Y'_{\dot{v}}$	-0.0407	-0.0504
$Y'_{r} - m'$	-0.0480	-0.0547
Y'_{vvv}	-1.5894	-1.7808
Y'_{vvr}	-0.4905	-0.6006
Y'_{vrr}	2.9578	1.4360
Y'_{rrr}	-9.9792	-9.9426
$N'_{\dot{v}} - m' x'_G$	-0.0003	-0.0003
$N'_{\dot{r}} - I'_z$	-0.0275	-0.0395

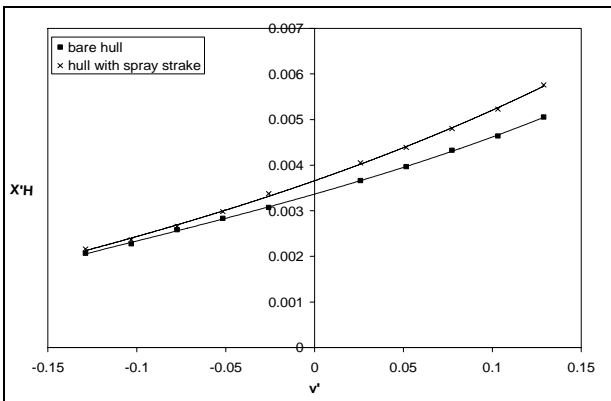


Figure 6: $X'H$ vs. v' graph

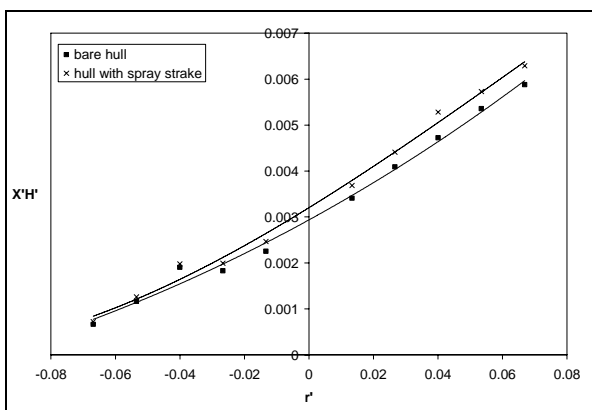


Figure 7: $X'H$ vs. r' graph

N'_{ν}	-0.0043	-0.0056
N'_r	-0.0011	-0.0017
$N'_{\nu\nu}$	-0.2574	-0.3510
$N'_{\nu r}$	0.0283	0.0788
N'_{rr}	-0.2672	-0.3677
N'_{rrr}	-0.9234	-1.1802

Table 4. Interactive coefficient

Interactive coefficients			
$1 - t_R$	0.8681	κ	0.4558
a_H	0.0226	l'_R	1.0194
ε	0.9018	γR	0.3799

Table 5 shows propeller diameters and rudder areas used in the simulation. Resistance predicted for planing hull can be obtained through model test in the towing tank. Propulsion prediction for use in simulation was calculated based on the Blount method in ref. [13]. The Propulsion choice is submerged propulsion type with 12 deg inclined shaft on propeller. Rudder areas are calculated based on the rudder area coefficient found in ref. [14]. It was found that the total rudder area of these craft can be expressed as rudder area = $0.0016L_p^2$

Table 5. Propeller and rudder parameters

Parameters	value
Ship Speed, V (knots)	15
Ship Resistance, R_T (kN)	42.28
Number of propellers	2
P/D	1.2
Number of Blades, Z	4
Diameter, D (m)	0.9
Blade Area ratio, EAR	100
RPS, n	18.28
Wake fraction, w (0.03 to - 0.05)	0.03
Trust deduction, t (0.07 to 0.11)	0.11
Advance ratio, Jp	0.63
Thrust coefficient, KT	0.27
Eff. Of Open water	0.595
Eff. Of Rotary rotative (0.97 ~ 1.01)	1.01
Eff. Of Hull	1.04
Eff. Of Propeller (0.62 ~ 0.65)	0.551
Thrust, T (kN)	47.0
Shaft angle (deg.)	12
Rudder Area, m ²	0.714

4. Discussion

Figure 10 shows the result of the simulation of the turning circle. From this simulation has been calculated that the tactical diameter of the vessel (bare hull) is 322.24 m, 16 times the vessel length, which is 20.19 m. This tactical

diameter is much more than the required IMO criterion, which is five times the vessel length. The advance is 354.25 m or 17.5 times the vessel length. This value is also much more than the IMO criterion 4.5 times the vessel length. Its result the same characteristics with Coccoli [2] and Soares [3] investigated on the manoeuvring by full scale sea trial on an HSC.

Figure 10 also shows the results of the hull with spray-strake, it has larger tactical diameter than the bare hull one. The reason may be the effect of configuration of spray-strake attaches. Through the design selection of number and position spray-strakes, the vessel can be maintain an improved manoeuvring performance and possibly increase its safety, such as the reduction of the tactical diameter in turning trajectory.

Figure 11 shows the result of the simulation for the 10-10 zigzag test. From this simulation, we can easily determine that the first overshoot is 4.250 deg, much less than the 10 deg of the IMO criteria. Figure 12 shows the result of the simulation for the 20-20 zigzag test. From this simulation also, we can see that the first overshoot is 4.007 deg, and the second overshoot is 3.920 deg. Both of these are much less than the value required by the IMO criteria. Thus, we decided for zigzag manoeuvre the vessel satisfies with IMO requirement

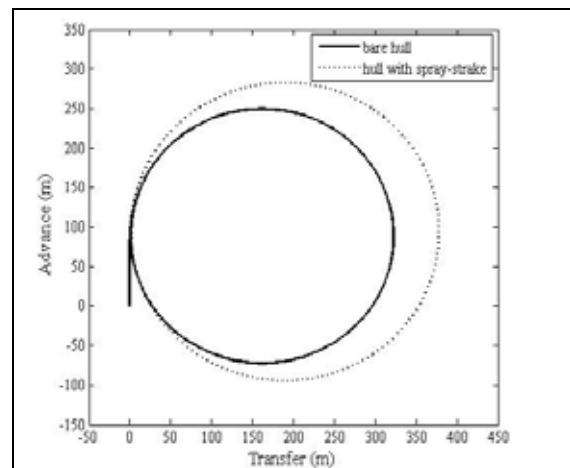


Figure 10: Comparison of 35 deg. of turning circle between bare hull and hull with spray-strake

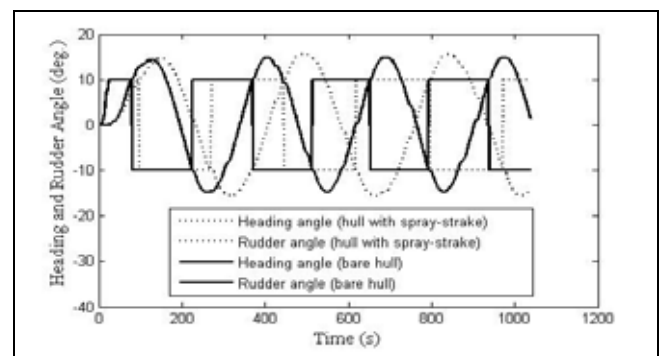


Figure 11 Comparison of 10/10 zigzag manoeuvre between bare hull and hull with spray-strake

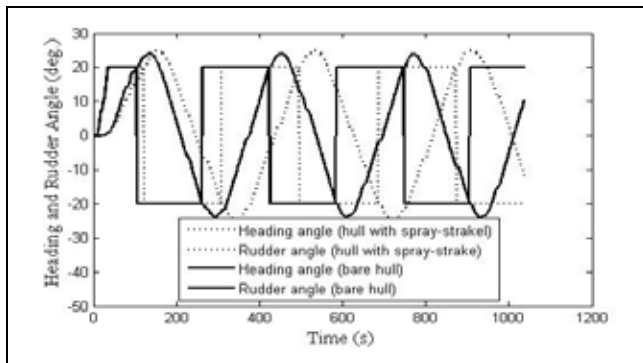


Figure 12 Comparison of 20/20 zigzag manoeuvre between bare hull and hull with spray-strake

5. Conclusion

Time domain simulation with the hydrodynamic coefficients derived from PMM tests developed in this research has shown significant effects of spray strakes attached on the planing vessel (patrol boat) in its manoeuvring performances.

Through the design selection of number and position spray-strakes, the vessel can be maintain an improved manoeuvring performance and possibly increase its safety, such as the reduction of the tactical diameter in turning trajectory.

Acknowledgments

The authors wish to thank IRPA, Ministry of Science Technology and the Environment for funding this work. Special thanks go to the friendly technicians of Marine Technology Laboratory of Universiti Teknologi Malaysia for their valuable assistance during the field measurement.

References

- [1] 23rd ITTC .2002. Manoeuvrability Evaluation and Documentation of HSMV” ITTC- Recommended Procedures, 7.5-02-05-05
- [2] Ikeda, Y., Katayama, T., and Okumura, H. 2000. Characteristics Of Hydrodynamic Derivatives in Maneuvering Equations for Super High-Speed Planing Hulls. *The Proceedings of The 10th International Offshore And Polar Engineering Conference*, Vol.4, Pp.434-444
- [3] Plante, M.C. 1998. Hydrodynamic Manoeuvring Aspects of Planing Craft. *Proceeding of DGA, International symposium and Workshop on Force Acting on a Manoeuvring Vessel*, Oct. 1998 Val de Reuil, France.
- [4] Coccoli, D., and Scamardella, A., 2004. High Speed Craft Manoeuvring Sea Trials. *Proceeding of 9th Symposium on Practical Design of Ships and Other Floating Structures* Luebeck-Travemuende, Germany.
- [5] Soares, C.G.,Francisco, R.A. and Laranjinha, M. 2004. Full-scale Measurements of the Maneuvering Capabilities of Fast Patrol Vessel, Argos Class, *Marine Technology*, Vol. 41. No 1, Jan 2004 pp 7-16.
- [6] Deakin, B 1998, Model Test to Assess the Manoeuvring of Planing Craft, *The International HISWA Symposium on Yacht Design and Yacht Construction*.
- [7] Abkorwiz, M.A. 1964. Lectures on Ship Hydrodynamics - Steering and Manoeuvrability. HyA Report HY-5, Hydro-Og Aerodynamisk Laboratorium, Lyngby, Denmark 1964
- [8] Lewis, eds 1989. Principles of Naval Architecture. volume 3:Trans. SNAME, Jersey City.
- [9] Sung, Y.J., Yum, D.J. and Rhee, K.P., 2000 An Analysis of The PMM Tests Using a System Identification Method, *The 7th International Marine Design Conference*, Kyung-Ju,Korea,May.
- [10]Condega, L. and Lewis, J. 1997 A case study of dynamic instability in a planing hull. *Journal of Marine Technology*. Vol.24, No.2, pp 143-163
- [11]Matora, S.,1959 On the Measurement of added mass and added moment of inertia for ship motion. (part 2. Added mass abstract for longitudinal motion) Do., Vol 106, (in Japanese)
- [12]Yoshimura, Y. and Ning Ma. 2003 Manoeuvring Prediction of Fishing Vessels, *Proceeding of Marsim 2003*, Japan
- [13]Blount, DL. 1997, Design of propulsion systems for high-speed craft, *Journal of Marine Technology and SNAME New*; Vol. 34, 4
- [14]Hadler, J.B. 1962. Prediction of the power Performance of the series 62 Planing Hull Forms, *Sname Transaction*

Feasibility Study of Ultrasonic Transducers as Altitude and Attitude Sensors for Flight Control System

T. Indriyanto^{1*} and **U. M. Zaeny**²

¹ Aerospace Engineering, Faculty of Industrial Technology
Institute of Technology Bandung, Bandung, Indonesia
Tel: +62-22-2504529, Fax: +62-22-2534164, E-mail: t.indriyanto@ae.itb.ac.id

² CV Ummah Sciences, Bandung, Indonesia
Tel: +62-22-2506968, E-mail: uuhmzst@yahoo.com

Abstract

Motion parameters of an aircraft need to be identified along its trajectory. In low-level flying region, such as take off and landing phases, measurement of altitude and attitude parameters are very important. However, standard (static-pressure based) altimeters for aircraft would not work properly at that region. While radar altimeter can be used for such purpose, a low-cost and light-weight device is often necessary. This paper describes the feasibility study of ultrasonic transducer for altitude and attitude sensors as one of postgraduate student researches conducted at Aerospace Engineering - ITB.

Investigation into the use of ultrasonic transducer was carried out by constructing a two degree-of-freedom fixed-based motion simulator. The simulator depicts an aircraft pitch attitude and elevator motion to stabilize it. Using the simulator, ultrasonic transducer was tested statically as low-level altimeter and attitude sensor, and dynamically as part of pitch attitude hold (PAH) control system. Research results showed some limitations and potential implementations of ultrasonic transducer as altimeter and attitude sensors.

Keywords:

Transducer, Altimeter and Attitude Sensors, Motion Simulator

1. Introduction

Motion parameters of an aircraft (e.g. airspeed, altitude, position and attitude in three-axes) need to be identified along its trajectory. In low-level flying region, such as take off and landing phases, measurement of altitude and attitude parameters is very important. In normal, cruising flight aircraft altitude can be measured using static-pressure altimeter, radar altimeter or GPS receiver while its attitude is obtained from gyroscopes and magnetic compasses. However, standard (static-pressure based) altimeters would not work properly at that region. While radar altimeter can be used for such purpose, a low-cost and light-weight device is often necessary, especially in the case of unmanned aerial vehicle (UAV). This paper describes the feasibility study of ultrasonic transducer for altitude and attitude sensors as one of postgraduate student researches conducted at Aerospace Engineering - ITB [1].

2. Approach and Methods

Investigation into the use of ultrasonic transducer was carried out by constructing a two degree-of-freedom

fixed-based motion simulator. The simulator represents an aircraft pitch attitude and elevator motion to control it. Using the simulator, ultrasonic transducer was tested statically as low-level altimeter and attitude sensor, and dynamically as part of a pitch attitude hold (PAH) flight control system.

2.1. Ultrasonic Transducer

Ultrasonic transducer is not a very special nor unique piece of instrument in today's life. Its wide applications include those found in physics labs, chemistry, imaging, medical and other modern technology equipment.

An ultrasonic transducer might have frequency between 20 kHz up to several GHz. For this research a 40 kHz transducer is chosen. In its basic form it costs about 5 USD each. A typical transducer is given in Figure 1.

The transducer has specifications as follows [2]:

Resonant frequency [kHz]	: 40
Sound pressure level [dB]	: 115 <
Sensitivity [dB]	: -64 <
Diameter [mm]	: 16.2
Height [mm]	: 12.2
Terminal interval [mm]	: 10.0

* Corresponding author

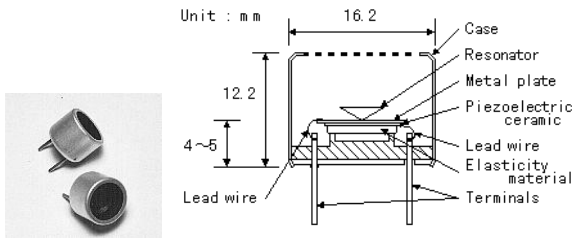


Figure 1. Typical ultrasonic transducer [2]

2.2. Ultrasonic Altimeter

To use ultrasonic transducer as altimeter, it has to be assembled with other electronic components to perform functions as shown in the following block diagram.

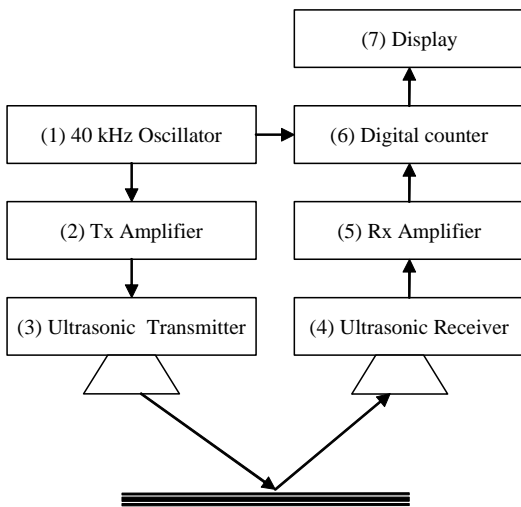


Figure 2. Block diagram of ultrasonic altimeter [3]

As shown in Figure 2, by measuring time Δt required for receiving the reflected signal, altitude can be calculated by multiplying Δt with the speed of sound and divided the result by two. At 20 deg C the speed of sound is approximately 343.5 m/s. The assembled components of an ultrasonic altimeter is given in Figure 3.

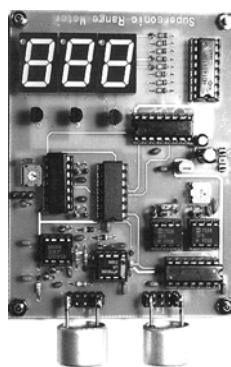


Figure 3. The circuit of an ultrasonic altimeter [1]

Several sets of altimeter were built to be used together as attitude sensor. The altimeters were tested and some of the results are given in Section 3.

2.3. Ultrasonic attitude sensor

Basically, for attitude sensor two sets of transducer need to be installed in aircraft: at nose and tail (for pitch angle measurement) or left and right wing (to measure roll angle). According to Figure 4 by measuring altitudes of the nose h_1 and tail h_2 , aircraft pitch angle θ can be determined.

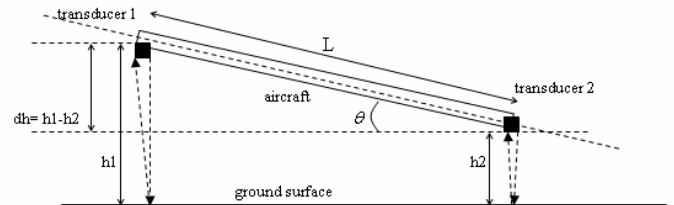


Figure 4. Principles of attitude measurement

To simplify the design of microcontroller used in flight control system which uses an 8 bit data format, standard 'arcus sinus' formula to determine θ was not used. An approximation formula was programmed into the microcontroller. It has been shown in [1] that for θ between -20 to +20 deg the formula in equation (1) can be used with errors in θ less than 2.5%. Since at low-level flight aircraft pitch attitude never exceeds that angle limits, the formula is considered to be acceptable.

$$\theta = \sin^{-1} \frac{dh}{L} \quad \frac{dh}{L} \quad (1)$$

However, to function correctly as depicted in Figure 4 the two transducers need to send signal almost perpendicular to the ground surface, i.e. both transducer should be rotated to follow the angle θ . This is a difficult design to built.

One solution to overcome this problem is by using two transducers at each tip of aircraft, meaning two transducers at nose and two at tail. Each transducer is fixed at a known angle as given in Figure 5. This angle will depend on available space in aircraft and the intended altitude to measure.

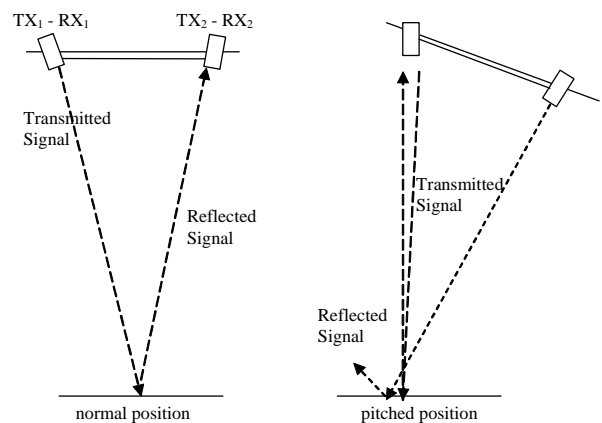


Figure 5. Two ultrasonic altitude sensors at aircraft nose

As illustrated in Figure 5, in normal position receiver RX₁ will receive signal from transmitter TX₂, and RX₂ signal from TX₁. The result is then averaged to obtain a value of altitude. In pitched position RX₁ will pick up signal from TX₁, while result from RX₂ will be ignored. Figure 6 shows the electronics of ultrasonic pitch sensor.

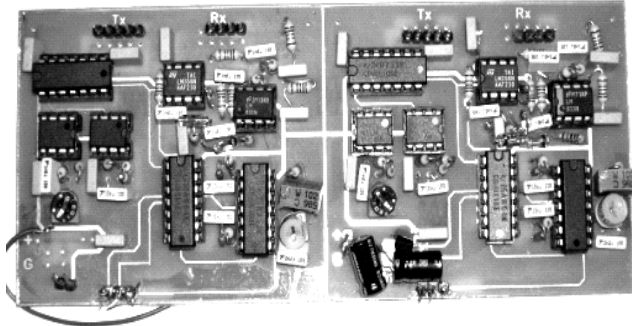


Figure 6. The electronics of ultrasonic pitch sensor [1]

2.4. Pitch Attitude Hold

In flight control a Pitch Attitude Hold (PAH) system is used to maintain aircraft pitch attitude at a predetermined position according to its phase of flight, e.g. -5 deg pitch down in descent phase or 2 deg pitch up in cruise phase. The block diagram of a PAH system is given in Figure 7.

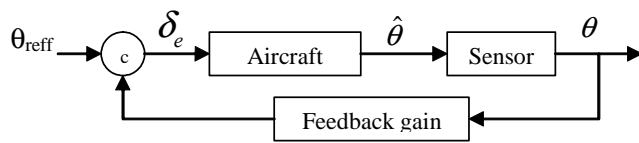


Figure 7. Block diagram of a PAH system

Input to a PAH system are the desired pitch angle θ_{ref} . Based on standard equations implemented in the system, PAH then generates command to change the elevator deflection δ_e to achieve the desired attitude.

In this research the PAH system was based on an 8-bit microcontroller AT89C51. The system should be able to receive input from a PC via serial port or the ultrasonic sensors. It should be able to produce command output to a servo controlling the elevator deflection. As shown in Figure 7 the algorithm for PAH is a simple feedback gain. The development of the circuitry was based on [4].

2.5. Motion Simulator

The motion simulator was designed and built to demonstrate the motions of pitch attitude (± 30 deg) and elevator control (± 20 deg). Figure 8 shows the block diagram of the motion simulator.

Aircraft equation of motion is modeled using state space equation [5]:

$$\frac{dx}{dt} = f(x,u) = A.x + B.u \tag{2}$$

where $x = \{\hat{u} \ \hat{a} \ \hat{\theta} \ \hat{q} \ dz\}^T$ and $u = \{\delta_e\}$.

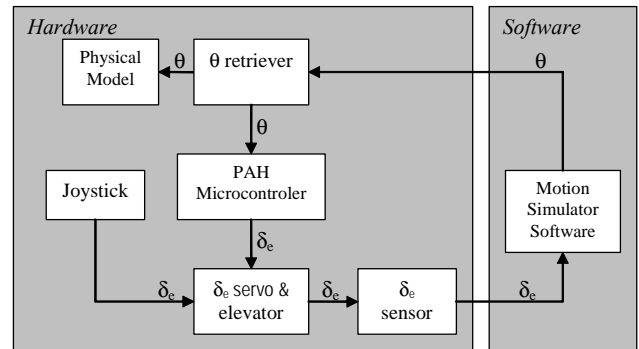


Figure 8. Block diagram of the motion simulator

The linear, time-invariant differential equation is solved using fourth-order Runge-Kutta method as follows [6]:

$$\begin{aligned} k_1 &= dt \ f(x,u) \\ k_2 &= dt \ f(x + \frac{1}{2}k_1, u) \\ k_3 &= dt \ f(x + \frac{1}{2}k_2, u) \\ k_4 &= dt \ f(x + k_3, u) \end{aligned}$$

and

$$x(t + dt) = x(t) + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \tag{3}$$

The physical model of the motion simulator and its elevator part is depicted in Figure 9.



Figure 9. The motion simulator [1]

2.6. Testing

Two kinds of testing were conducted. Static tests were conducted to investigate the use of ultrasonic transducers as

low-level altimeter and attitude sensor. In these tests the simulator was fixed at certain altitude and pitch attitude to allow the measurement by the transducers.

Dynamic tests were conducted to investigate the ability of the transducers to work together as part of PAH system in realtime. In these tests the simulator was given initial condition and the PAH tried to stabilize it at a predetermined pitch attitude. Results of both tests are given in the following section.

Before dynamic tests were performed, extensive tests were also conducted to ensure that the motion simulator and its components work as designed.

3. Results

Result of static tests are given in Figures 10, 11 and 12.

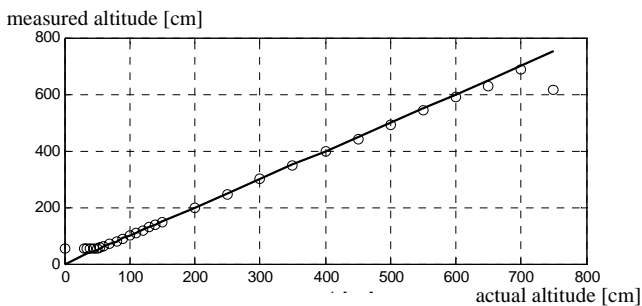


Figure 10. Actual and measured altitude

Figure 10 shows the actual altitude compared with those measured by the ultrasonic altimeter. From this result, it is clear that the ultrasonic transducer is capable of measuring altitude between 55 - 700 cm quite accurately. The measurement accuracy depends on the electronics used (e.g. LED display and number of bit used) and in this case it is accurate to 1 cm. In average the error in measured altitude varies from 1 - 5%.

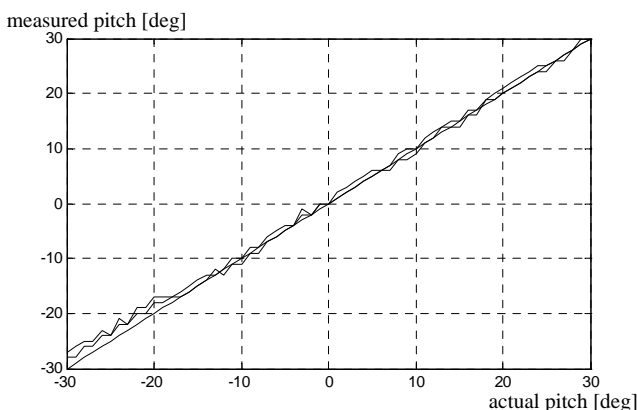


Figure 11. Actual and measured pitch attitude

Figure 11 shows the actual attitude compared with those measured by the ultrasonic pitch sensor. It can be seen that measuring range between -30 to +30 deg could be achieved. The measurement is accurate of 1 deg, but this value is

limited by electronics used. The error also depends on attitude itself and averaged to 1 deg. Maximum error of 3 deg occurs when pitch angle reaches its minimum at -20 deg.

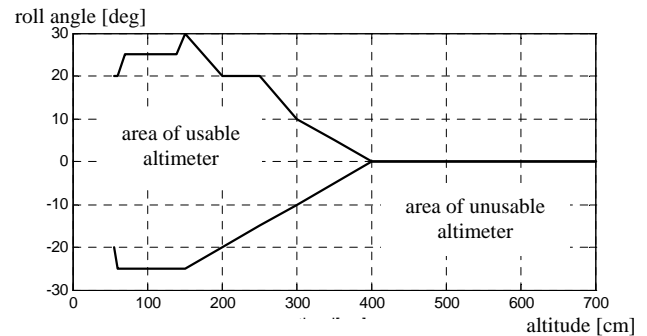


Figure 12. Effect of altitude to maximum roll angle

Figure 12 shows the effect of altitude to the allowable (maximum) roll attitude at which the altitude sensor can be used. It is clear that maximum roll angle depends on altitude and varies from ± 20 deg at 55 cm to zero at 400 cm. To measure altitude higher than 400 cm the roll angle must be close to zero.

Result of dynamic tests are given in Figures 13 and 14.

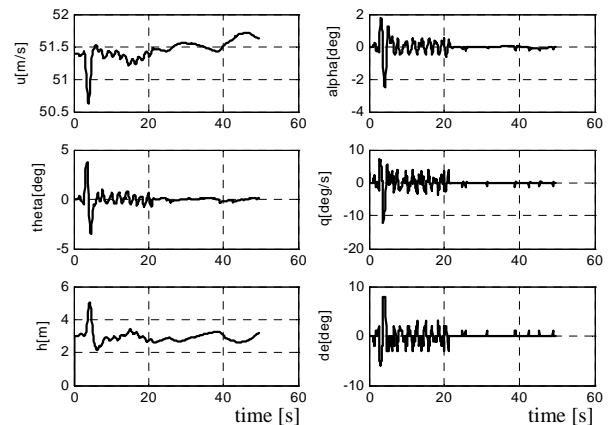


Figure 13. Closed loop simulation of PAH system with feedback gain 0.75

Figure 13 shows the time histories of airspeed u , angle of attack α , pitch angle θ , pitch rate q , altitude h and elevator deflection δ_e . The intended (initial) altitude is 3 m with given disturbance in pitch angle 4 deg. Using feedback gain 0.75 it is clear that the PAH system works very well by deflecting the elevator to maintain the altitude and eliminate pitch angle and pitch rate to 0 deg. During simulation pitch angle of less than ± 1 deg could be maintained.

Figure 14 shows the same conditions with feedback gain of 3.0. It is clear that the results are not as good as previous ones. The altitude deviates very large and the elevator is deflected up and down to its limits. The feedback gain of 0.75 was found to be the best for the PAH control system.

In addition, test results show the capability of ultrasonic transducers to support data processing speed of 10 Hz. The physical model is also capable of demonstrating motions of

pitching between -35 to $+35$ deg and elevator control between -20 to $+20$ deg. Both motions were performed within 1 deg accuracy.

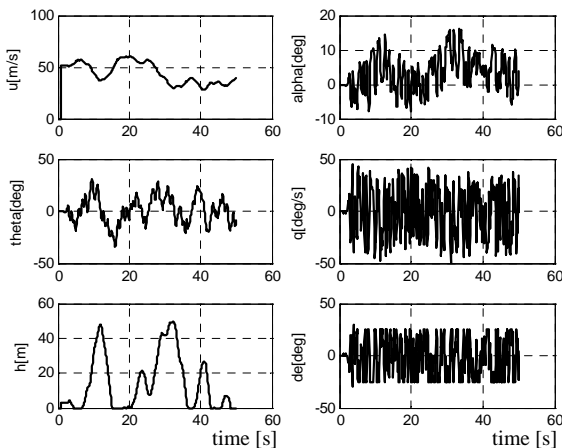


Figure 14. Closed loop simulation of PAH system with feedback gain 3.0

4. Discussion

Ultrasonic transducer has shown its capability to be used as altitude and attitude indicators, and as part of a control system. It can be further developed and installed on small aircraft requiring more precise measurements in addition to those provided by static pressure port, GPS or gyroscopes. In this research four transducers were placed on the nose and tail to measure aircraft pitch attitude. Similar configuration can be installed on aircraft wing to enable the measurement of roll angle.

The accuracy and range of measurement can be enhanced by developing more complex electronics. By reducing the period of transmitted signal and pulse width, altitude measurement of less than 50 cm can be achieved, and vice versa for measurement of more than 700 cm. Filters can also be integrated into the system to reduce noise and enhance its accuracy. More accurate angle measurement can be achieved by separating the two transducers at longer distance, but this method will depend on available space in aircraft and the intended altitude to be measured.

5. Conclusions

The feasibility study of ultrasonic transducers as altitude and attitude sensors for flight control system has been performed in this study. Ultrasonic altimeter, pitch attitude sensor and motion simulator were built and tested.

The altimeter has the capability of measuring altitude between 55 to 700 cm, accurate to 1 cm, and has error of 2% in average. It has to be noted, however, that roll angle should be kept less than 20 deg to guarantee the working condition of the altimeter.

Ultrasonic attitude sensor has the capability of measuring pitch angle between -20 to $+20$ deg, accurate to 1 deg, and has error of 1 deg in average. It has to be noted, however, that altitude will affect on the working condition of the attitude sensor.

The use of ultrasonic transducers in control system was also demonstrated. It was found that the optimum feedback gain was 0.75. The gain has very large effect on the overall performance of the PAH system. Data processing speed of 1 Hz can be achieved. This speed can be increased further by improving the electronics as required by the control system.

The above results show some limitations and potential implementations of ultrasonic transducer as altimeter and attitude sensors. Further development include the incorporation of these sensors in an unmanned aerial vehicle (UAV) flight control system to enable the automatic landing of the aircraft.

6. References

- [1] Zaeny, U. M. 2006. Construction and Testing of Ultrasonic Sensor as Feedback for Pitch Attitude Hold Automatic Control System (in Indonesian). MSc Thesis, Aerospace Engineering, Institute of Technology Bandung.
- [2] Anon. 18 March 2003. Citing Internet sources URL http://www.interq.or.jp/japan/se-inoue/e_srm.htm
- [3] Syaefullah, M. N. 2003. *Design and Implementation of Proximity Meter Based on Microcontroller AT89C2051* (in Indonesian). BSc thesis, Dept. of Physics, Faculty of Mathematics and Pure Sciences, Institute of Technology Bandung.
- [4] Putro, A. E. 2000. *Learning Microcontroller AT89C51/52/55 Theory and Application* (in Indonesian), Yogyakarta, Gava Media.
- [5] Jenie, S. D. 1991. *Lecture Notes on Flight Control* (in Indonesian), Dept. of Aerospace Engineering, Institute of Technology Bandung.
- [6] Kreyzig, E. 1999. *Advanced Engineering Mathematics*. 8th Edition, John Wiley & Son.

Development of Cavity Defect in Forward Extrusion

Tri Widodo Besar Riyadi¹, John Atkinson², Waluyo Adi Siswanto^{1*}

¹Department of Mechanical Engineering
Muhammadiyah University of Surakarta, Pabelan, Surakarta, Indonesia.
Tel: +62-271-717417 ext 222, Fax. +62-271-715448, E-mail: tri_wbr@yahoo.com

²School of Mechanical Aerospace and Civil Engineering,
The University of Manchester, Oxford Road, Manchester UK M13 9PL
Tel: +44 (0) 161 306 6000

Abstract

Cavity defect occurs in an extruded product when the remaining billet reaches a certain thickness. This defect can create serious problem in the quality of products because it produces waste material and requires an extra operation for finishing. The requirement of wasting a certain portion of the product becomes undesirable especially for a relatively short, headed product. Since the occurrence of this defect is troublesome in industrial practice, it is important to study the conditions leading to the cavity formation. By using this prediction, it may be possible to stop the process before the cavity starts to develop.

In this work, a new measuring technique to monitor the development of cavity defect in forward extrusion was studied. This technique uses electrical contact principle, which detects the contact between metal pins and the leading face of the billet. When the cavity develops, a void or gap between the punch and the billet interface develops, and breaks the electrical current of the circuit. The experiment of extrusion process was undertaken in parallel with finite element analysis (FEA) using ABAQUS software. It was shown that at a certain distance the cavity defect was revealed. The cavity formation was investigated by a variation of die reduction and the influence of coefficient of friction, punch speed, and billet thickness.

This study concentrated on improving the understanding of the behavior of metal flow and the development mechanism of cavity formation with the aim of providing general guidance for the extrusion industry in minimizing the cavity defect formation of the extruded product.

Keywords:

Cavity Defect, Forward Extrusion, Measuring Technique, Finite Element Analysis

1. Introduction

The goal of any extrusion manufacturer is to produce parts that not only conform to dimensional and metallurgical specifications but also have a high product quality or with a minimum number of defects. There are various types of defect observed on the surface of, and inside, extrusion products. Johnson and Kudo (1962) classified the defects in extrusion into three groups: cracking, sinking-in or contracting, and skin inclusion. Figure 1 illustrates an extrusion defect called sinking-in which happens when a part of billet surface sinks, thus forming a recess or cavity. Cavity defect as shown in Figure 1 occurs in an extruded product when the remaining billet reaches a certain thickness.

Since the occurrence of this defect is troublesome in industrial practice, it is important to study the conditions leading to the cavity formation. The study of cavity is also interesting for understanding the flow pattern of material during deformation process. Furthermore, knowledge

about the cavity characterization is essential for the design of an effective process.

The study of cavity can be found in the works by some investigators. The first report of the cavity phenomenon was that of Genders [1] who studied the enfolding of impure skin into a slug during extrusion. This defect, which is commonly called "piping", appeared as a tubular accumulation of impurities lying co-axial in the slug. Johnson [2] studied the cavity formation using a profiled punch to facilitate the flow pattern. According to Johnson, there is a relationship between the cavity formation and the pressure load. It was found that, when the coring point happened, pressure load decreased rapidly. In 1961, Johnson and Mellor [3] showed that the coring phenomenon happened when the slug is about half the diameter of the extruded rod. An expression was then proposed to compute the cavity size at any slug length.. Avitzur [4] analyzed the process of axi-symmetrical extrusion through a chamber using an upper bound technique. This technique has been used to predict the

occurrence of cavity defect in axi-symmetrical billet during the extrusion process. Atkinson [5] also investigated the coring defect during forward extrusion. Concave and convex punches were used in this observation to examine the effect of punch profile in the plane strain billet on the start of cavity formation. Atkinson also used an interruption technique to investigate the development of cavity defect in a cylindrical billet. Different remaining thickness and cavity shapes were obtained from stopping the punch displacement at different positions. Recio [6] investigated the cavity formation in forward extrusion using flat specimens with different slug lengths. It was found that the development of cavity size depended on that residual slug length and the die orifice. Riyadi [7] developed a measuring technique to monitor the development of cavity defect. It was found that the cavity develops nearly linearly.

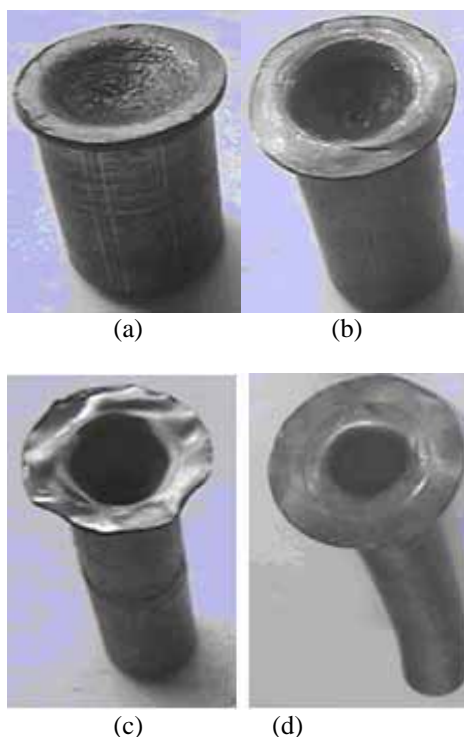


Figure 1. Cavity Defect Formed in the Back End of Extruded Billets

In this paper the influence of a variation of die reduction to the formation of cavity is studied. The study is connected using the measuring technique developed by Riyadi. A finite element simulation is also carried out using ABAQUS [8] in order to compare the experimental result and the numerical result. By using simulation method, the experiment has also allowed to study the effect of friction coefficient, billet thickness and punch speed. Meanwhile physical modeling was outstanding to give information about the flow pattern of material.

2. Experimental Methods

There is a requirement to re-build a special punch, which combines a hollow punch and a special pressure pad with some holes inside. Metal pins that are connected with

insulated cables are attached in the hole of pressure pad. The head of the pins on the leading face of the pressure pad will contact to the back face of the billet as the punch presses on the billet. Pins are connected to a power supply and the billet is linked to a body ground. Electrical circuit is then built between the pins and the billet. In order to get electrical signal, the billet material must be conductive. Lead is chosen as billet material due to its electrical conductivity and less expensive. The concept of the measurement is that there will be electrical contact between pin contact at the leading face of the punch and the back face of the billet before cavity formation. When the cavity starts or develops; a void or gap between the punch and the billet interface develops and the contact would become uncovered. This gap breaks the electrical current and thus the resistance of the circuit becomes overloaded. Data acquisition will record the resistance-displacement data and which will then be displayed on the computer screen within a very short time during the process.

Prior to commencement of the tests, all equipments such as the container chamber, the die, the punch and the specimen were cleaned and lubricated by grease. Lubrication was given so that the punch is free to move inside the container. This is also to minimize the friction effect in the billet-container and the billet-die interfaces. An excessive amount of friction may increase the punch load and may give effect to skin inclusion defect.

Extrusion tests were carried out in an Instron machine with maximum capacity of 200 kN. Punch displacement was set equal to the length of the billet for each die. Using this length, it is expected that the extrusion would produce completely extruded product. The first result obtained from this test is given in the form of punch load versus punch displacement data. The application of present measuring technique has then allowed obtaining resistance measurement during the process. The resistance measurement was designed to determine the punch displacement at the instant when the billet has lost contact with the pin. The sequence of the pins which have lost contact then reveals the development of the cavity at the leading face of the billet. Before cavity happened, all pin contacts were connected to the billet face. When lost-contact occurred, the magnitude of the resistance detected by means of five pins would be very high. The sequence of the lost-contact of the pins happens since each pin has different radial positions.

The measurement includes the propagation of the cavity diameter from its initiation until the extrusion process is interrupted. During the measurement, it is defined that the measurement must not affect the loading process in any way and the extrusion process should allow continuous measurement to enable the study of the cavity development.

The simulation model using ABAQUS is solved by using the use of adaptive meshing in order to avoid a large distortion in the billet during the process. The model is axi-symmetric and consists of two rigid tools (punch and die) and a deformable billet. The rigid tools are modeled as analytical rigid surfaces. In the first experiment, all contact

surfaces are assumed to be well-lubricated and, thus, are treated as frictionless. A penalty is given to change the magnitude of friction coefficient. The billet is made of lead and is modeled as a von Mises elastic-plastic material with isotropic hardening. Material data is obtained from compression testing using a cylindrical specimen with 25.4 mm diameter and 19.0 mm thickness. Spiral groove was given in both bottom and upper surfaces in order to facilitate lubrication which used grease. Lubrication is critical in compression test in order to avoid barreling. The Young's modulus is $6.25 \cdot 10^8$ Pa, and the initial yield stress is 8 MPa. The Poisson's ratio is 0.43; and the density is $11,400 \text{ kg/m}^3$.

In the first step the punch is moved to a position where a contact between punch and billet is established and slipping of the workpiece against the die begins. In the second step the punch is extruded through the die to realize the extrusion process. This is accomplished by prescribing displacements to the nodes at the top of the punch.

3. Results and Discussions

3.1. Material Flow Pattern

The cavity formation in extruded billets might be analyzed by the obtained flow pattern of the billet after deformation. The flow pattern of material can be seen from the distortion of plastically deformed region, which also shows the distribution of strain. Figure 2 shows the experimental flow pattern of extruded materials obtained from physical modeling and prediction by finite element simulation in a frictionless condition. For all reductions, agreement between the actual flow pattern and that predicted by the finite element method is generally good. In general terms, results of physical modeling and finite element simulation are in good agreement in terms of magnitude of the strains and their distribution, although some discrepancies are found to occur towards the exit of the plastically deforming region.

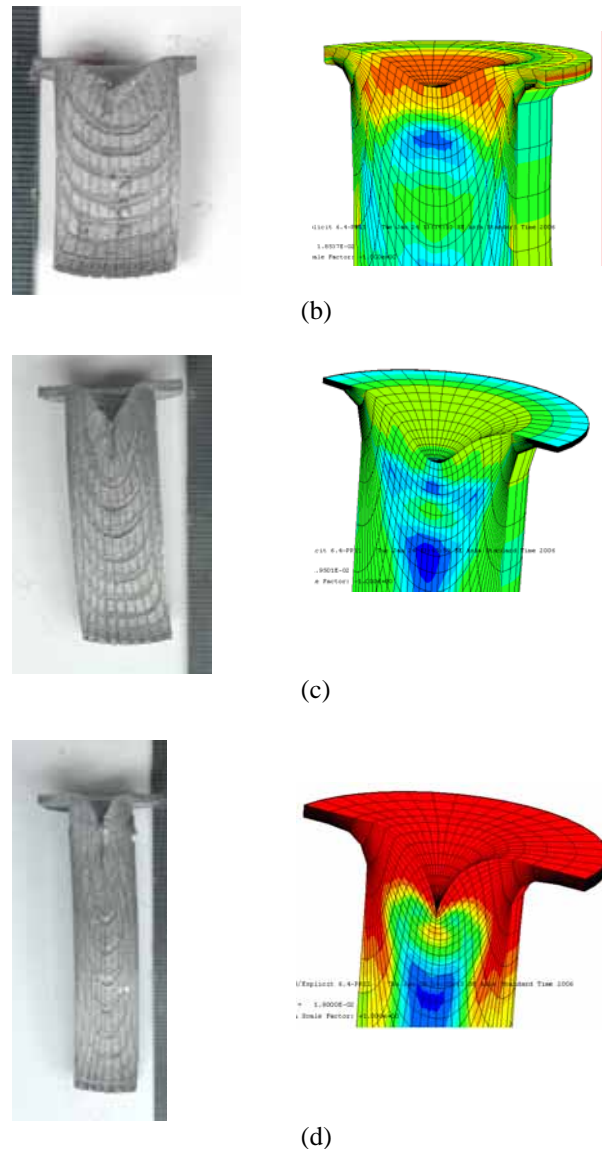
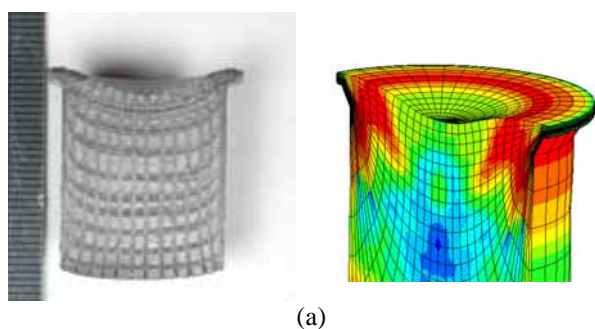


Figure 2. Flow Pattern of Material Obtained By Experiment and Simulation
(a). $R = 0.25$, (b). $R = 0.44$, (c). $R = 0.60$, (d). $R = 0.74$

3.2. The Onset of Cavity Formation

The onset of cavity formation is a term used to indicate the initiation point when the cavity starts to develop. The thickness of the billet when the cavity starts to develop is called the critical thickness. Information of the critical thickness of the billet in extrusion is important to enable the process could be stopped before the commencement of cavity formation. In the present experiment, the onset of cavity formation of an extruded billet has been investigated by using a new measuring technique developed by Riyadi, which used electrical pins as electrical current detector. The setting up of the equipment is described Figure 7 in the end part of this paper.

Figure 3 shows the electrical resistance levels detected by a means of five pins during the extrusion process plotted against punch displacement. It can be seen that the

resistance levels are different for each of different pins. The resistance of the pins fluctuates and stops at different punch displacements. Fluctuation shows that the resistance of a material is influenced by the distance between the pin and the billet body. The magnitude of the pin resistance in the present work is not of concern. The required information to monitor the development of the cavity is the loss of contact between the pin and the billet which is showed by the end line of the resistance. The loss of contact between the pin and the billet shows that there is a void or gap between the pin and the billet. The first lost contact is referred to as the onset of cavity formation.

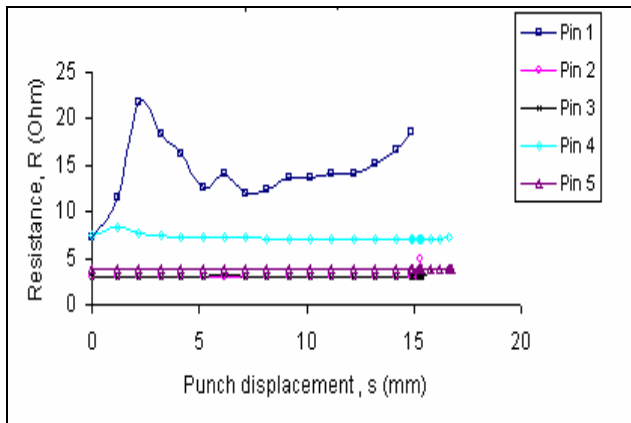


Figure 3. Resistance Detected By 5 Metal Pins

The critical thickness which shows the onset of cavity formation for all reductions can be more clearly seen in Fig 4. It can be seen that the critical thickness decreases rapidly as the reduction increases. This means that the onset of the cavity formation is greatly influenced by reduction. It can also be said that an increase in reduction tends to delay the onset of cavity formation.

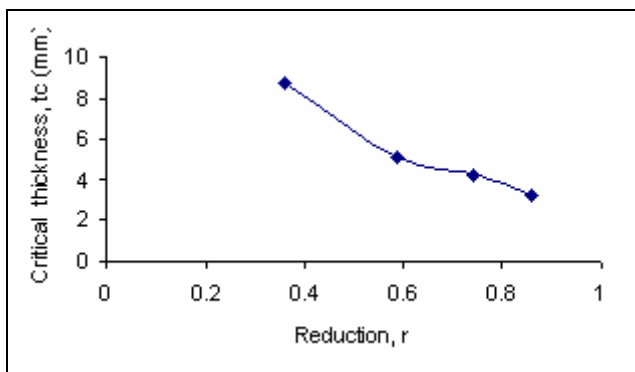


Figure 4. The Critical Thickness for Different Reductions

3.3. The Effect of the Coefficient of Friction, Punch Speed and Billet Thickness on the Cavity Formation

The experiment to investigate the effect of the coefficient of friction on the cavity formation was carried out by using different values of the coefficient of friction. The result of experiment is shown in Figure 5. It can be seen that the increase of the coefficient of friction tends to postpone the cavity formation. An increase of the friction

coefficient also causes the increase of punch force required to perform the extrusion process.

The experiment to study the influence of the punch speed on the cavity formation was carried out by using different values of punch speed. Figure 6 shows the result obtained in the present simulation which shows that an increase in the punch speed has produced smaller cavity. This result has proved a theory which states that an increase in strain rate will cause a material to be stronger.

Other experiment which was carried out by simulation has also been successful to study the influence of billet thickness on the cavity formation. The result of this experiment shows that the change in the billet thickness has no significantly effect on the cavity formation.

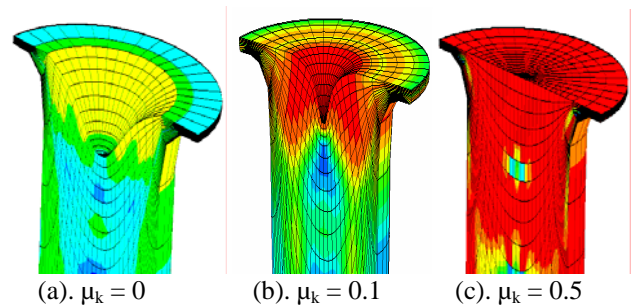


Figure 5. The Coefficient of Friction Effect on the Cavity Formation

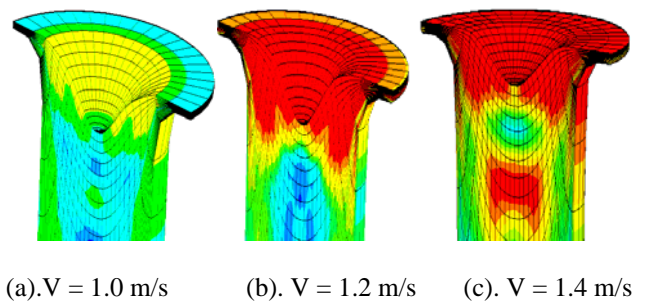


Figure 6. Punch Speed Effect on the Cavity Formation

4. Conclusions

The present investigation has performed a forward extrusion test, which was carried out at cold temperature using lead specimens through a square die. From the data obtained and the analysis based on the present experiment, it has been shown that the electronic sensing can be used to monitor the development of the cavity defect formation in forward extrusion. The effect of reduction on the onset of cavity formation has also been investigated. As reduction increases, the onset of cavity formation decreases. The critical thickness therefore decreases as reduction increases. Other experiment which was carried out by simulation has also shown the influence of the coefficient of friction, punch speed and billet thickness.

Acknowledgements

The authors would like to thank TPSDP UMS for their support of this project.

References

- [1]. Genders, R. 1921. The Extrusion Defect. *J. Inst. Metals*, v.26, No.2, pp. 237-259.
- [2]. Johnson, W. 1955. Further Experiments in Plane Strain Extrusion. B.I.S.R.A. Reports MW/E/49/55.
- [3]. Johnson, W. and Mellor, P. B., 1983. *Engineering Plasticity*. Ellis Horwood Ltd.
- [4]. B. Avitzur, 1965. Analysis of Metal Extrusion. *Journal of Engineering for Industry, Trans, ASME, Series B*, vol. 87, pp.57-70
- [5]. Atkinson, J. 1972. An Investigation of the Coring Defect during Forward Extrusion of Lead. M.Sc Dissertation, Mechanical Engineering Department, UMIST, UK.
- [6]. Recio, E.S. 1993. Computer Simulation of The Forward Extrusion Process Using a CAD package. M.Sc Dissertation, Mechanical Engineering Department, UMIST, UK.
- [7]. Riyadi, TWB. 2004. Characterisation of Cavity Defect in Forward Extrusion, MSc Dissertation, Department of Mechanical Aerospace and Manufacturing Engineering, UMIST, UK.
- [8]. ABAQUS 65 SE User's Manual. 2004. ABAQUS, Inc.

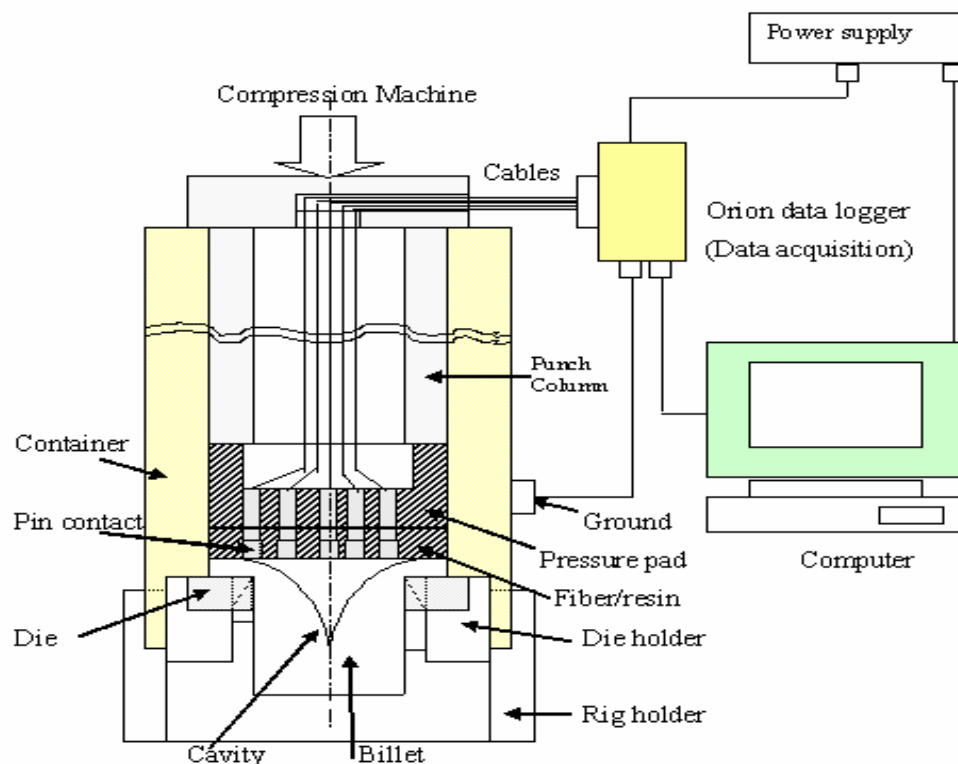


Figure7. A. Schematics of the Extrusion Equipment with Electrical Circuit Systems

Numerical Simulation of Jet Impingement Cooling on a Smooth Concave Surface

Suzairin bin Md Seri and Vijay R. Raghavan*

Faculty of Mechanical and Manufacturing Engineering
Kolej Universiti Teknologi Tun Hussein Onn(KUiTTHO)
86400 Parit Raja, Batu Pahat, Johor, Malaysia
Tel: +60-7-4537794, Fax: +60-7-4536080, E-mail: suzairin@kuittho.edu.my

Abstract

Jet impingement has been widely used as a means of heat removal because of its advantages in effective removal of locally concentrated heat and easy adjustment to the location where cooling is needed. Typical applications are paper drying process, electronics cooling, annealing of glass and elimination of excessive thermal load near the leading edge of gas turbine blade inner surface. More studies of jet impingement cooling are reported on flat surfaces than on concave and convex surfaces. For the flows on concave surface, the centripetal force due to the curvature makes the flow unstable and produces Taylor-Gortler vortices. Such vortices are known to enhance momentum and energy transfer and thereby heat transfer rate on the surface. The present study involves the simulation of a single slot jet impinging normally on a smooth concave surface by means of the Computational Fluid Dynamics software FLUENT. The velocity decay at the jet-center-line is simulated and compared with the experimental data of Choi et al. Two dimensional domains and steady state equations with Reynolds stress model for turbulence modeling were used for all simulations and good agreement between simulated results and the experimental data has been obtained.

Keywords: jet impingement, concave surfaces, enhancement, numerical simulation

1. Introduction

Rapid heating, cooling or drying of a variety of industrial products with large surface area is often provided using arrays of round or slot impinging jets. A single jet is used when localised heating or cooling is required and an array of jets is used for heating over a large surface. Heat transfer rates obtainable with impinging air jets are an order of magnitude higher than those usually associated with gaseous heat transfer media. Applications of impinging air jets include drying of paper, film and textiles, annealing of metal, glass and plastic sheets, cooling of electronic equipment etc.

Due to their common use in industrial heat and mass transfer applications, impinging jets have been studied by a number of authors both experimentally and numerically. In one of the first comprehensive studies of its kind, Gardon and Akfirat [1] obtained experimental heat transfer data for a single, submerged slot jet impinging on a flat surface. Their experiment spanned a range of jet flow speeds from an initially-laminar jet with a Reynolds number of 450 to a turbulent jet with a Reynolds number of 20,000 with nozzle-to-plate distances ranging from 2 to 32 jet-widths. The impingement plate was maintained at a constant temperature. Earlier they also investigated the effect of

coefficients for impinging slot jets. It has been concluded that the heat transfer characteristics could not be explained without accounting for the influence of turbulence, which again appeared to be dependent on the jet inlet conditions such as the jet inlet Reynolds number and jet-width. Therefore it is important to specify the proper jet-inlet conditions in terms of turbulence parameters and Reynolds number when trying to simulate jet impingement problems.

Nakatogawa and Nishiwaki [2] studied experimentally the impingement of a round jet on a flat surface. They too confirmed the strong relation of turbulence to the heat transfer values. Russell et al. [3] predicted numerically the axial velocity decay and pressure distribution along a flat surface for slot jet impingement, for Re of 11000 and 43000 and H/B of 15 to 30. H is the distance between the nozzle exit and the stagnation point of target and B is the nozzle width. Sparrow and Wong [4] investigated experimentally the impingement heat transfer coefficients on a flat surface for initially laminar slot jets. It has been concluded that the distribution of heat transfer coefficients was bell-shaped and that the shape of the initial velocity profile has a significant effect on the heat transfer from the impingement surface. For accurate prediction of impingement heat transfer coefficients on flat surfaces, the inlet velocity profile has to be input as accurately as possible.

Martin et al. [5] reviewed the earlier works on impingement that included single round jets and slot jets as well as arrays of round jets and slot jets. Correlations were given for average Sherwood numbers in terms of Re for many cases. Looney et al. [6] predicted lateral velocity

* Corresponding Author. E-mail: vijay@kuittho.edu.my,
Tel: +60-7-4537794, Fax: +60-7-4536080
turbulence on the distribution of local heat transfer

profiles at wall-jet region and skin friction distribution along the impingement surface for a slot jet for Re of 22000 to 42000 with H/B varying from 15 to 30. Polat et al. [7] reviewed comprehensively the earlier numerical studies on flat-plate-impingement. They found that most of them used the k- ϵ model and wall functions for near-wall treatment. The results were not satisfactory, especially the distribution of local heat transfer coefficients along the impingement surface. They attributed this to two things, one the turbulence model and the other, the wall functions. They questioned the validity of wall function in the impingement region arguing that the flow characteristics in this region significantly deviated from the ordinary flows over walls.

Cooper et al. [8] and Craft et al. [9] used the jet impingement problem as a test case for validating turbulence models. They investigated extensively the performance of k- ϵ model and second moment closure model for round jet impingement and concluded in favour of the Reynolds Stress Model for jet-impingement studies. Huang et al. [10] studied experimentally a single round jet for the distribution of local Nusselt numbers for a Re of 6000 and a wide range of nozzle-to-plate distances. Chen and Modi [11] investigated numerically the mass transfer coefficients of a turbulent slot jet impinging normally on a flat plate. They studied a wide range of Re from 450 to 20000 with H/B of 2 to 8. Lee et al. [12] studied heat transfer from a convex hemispherical surface to a round impinging jet. Chung et al. [13] investigated experimentally, the heat-transfer characteristics of an axisymmetric jet impinging on the rib-roughened convex surface.

Yang et al. [14] studied the effect of jet flows from three different slot nozzles with different nozzle-exit conditions on the impingement heat transfer characteristics on a semicircular concave surface. They studied the effect of Reynolds number, surface curvature and nozzle-to-plate spacing on the heat transfer characteristics. Their concave surface had a diameter of 30 jet-widths. They concluded that the effect of curvature was to increase the average heat transfer coefficient even though the stagnation point heat transfer does not change much. They also found that the effect of curvature becomes more prominent as Reynolds number increases. Choi et al. [15] continuing the work of Yang et al. produced experimental data for centre-line velocity decay along with the distribution of local heat transfer coefficients for Re ranging from 1780 to 4740. They also measured wall-jet profiles for low nozzle-to-plate spacing for a Reynolds number of 2960.

2. Approach and Methods

2.1. Case Description

A numerical study has been carried out for jet impingement cooling of a single slot jet impinging normally on a smooth semi-cylindrical concave surface, which diameter is 150mm. The jet fluid is the same as the stagnant fluid into which the jet emerges. Nine cases of jet-impingement on a concave surface, on which experimental

data are available, have been simulated. These cases involved the jet-inlet Reynolds numbers, Re_{2B} , of 4740, 2960 and 1780 with the nozzle-to-target distance, H/B, of 10, 6 and 4 jet-widths. The characteristic length for the Reynolds number is the nozzle exit hydraulic diameter which is 2B. The velocity decay at the jet-center-line is simulated and compared with the experimental data.

2.2. Method used

The two-dimensional computational domain is shown in Figure 1. Since the shape of the nozzle wall also could have a significant influence on the entrainment and flow thereon, its geometry was reproduced as nearly as possible as the one used by Choi et al [15] in their experiment. The nozzle geometry is shown in Figure 2.

In the present study, FLUENT, commercial CFD software is used for the simulations. This software is based on the finite-volume method. Most researchers dealing with impinging jets recommend the finite-volume method claiming good accuracy.

Two-dimensional domains are used and steady state solver is used for all the simulations. Radiation is neglected, as the temperatures involved are low. The segregated implicit solver is used because it is less time-consuming than the coupled implicit scheme for low-velocity flows. The SIMPLE scheme is used for pressure-velocity coupling. The QUICK scheme for discretization is used for the mass, momentum and energy balance equations. For turbulence modeling, the Reynolds Stress Model was used. The k- ϵ and k- ω model were also tried. The near-wall region is resolved all the way down to the wall and wall-functions are neglected since the validity of the wall-function approach in the impingement region was found to be questionable by earlier researchers (Polat et al [7]) in the area.

2.3. The Governing Equations

2.3.1. Continuity Equation

$$\frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} = 0 \quad (1)$$

2.3.2. Momentum Equations

$$u \frac{\partial(\rho u)}{\partial x} + v \frac{\partial(\rho u)}{\partial y} = -\frac{\partial p}{\partial x} + \frac{\partial(\tau_{xx})}{\partial x} + \frac{\partial(\tau_{xy})}{\partial y} + \frac{\partial(-\rho \overline{u^2})}{\partial x} + \frac{\partial(-\rho \overline{u'v'})}{\partial y} \quad (2)$$

$$u \frac{\partial(\rho v)}{\partial x} + v \frac{\partial(\rho v)}{\partial y} = -\frac{\partial p}{\partial y} + \frac{\partial(\tau_{yy})}{\partial y} + \frac{\partial(\tau_{xy})}{\partial x} + \frac{\partial(-\rho \overline{v^2})}{\partial y} + \frac{\partial(-\rho \overline{u'v'})}{\partial x} \quad (3)$$

$$\text{where } \tau_{xx} = 2\mu \frac{\partial u}{\partial x} - \frac{2}{3}\mu \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right)$$

$$\tau_{yy} = 2\mu \frac{\partial v}{\partial y} - \frac{2}{3}\mu \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right)$$

$$\tau_{xy} = \tau_{yx} = \mu \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right)$$

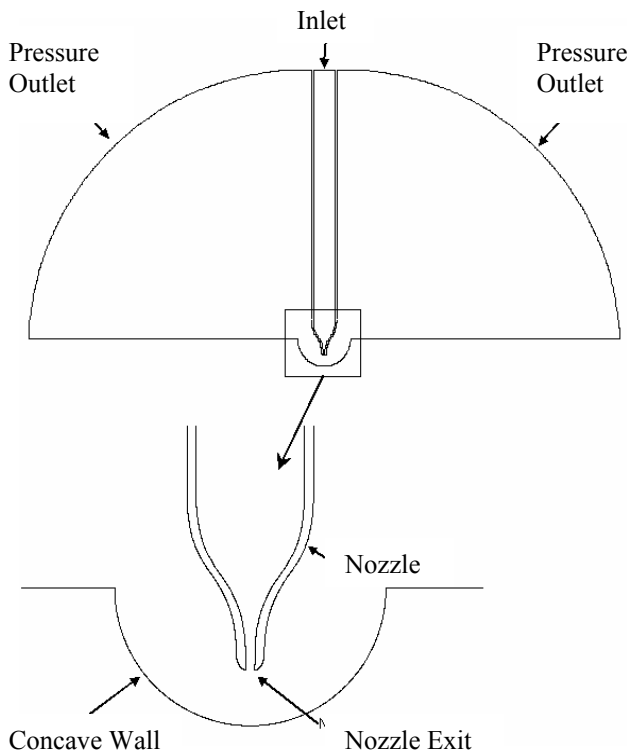


Figure 1: Computational Domain

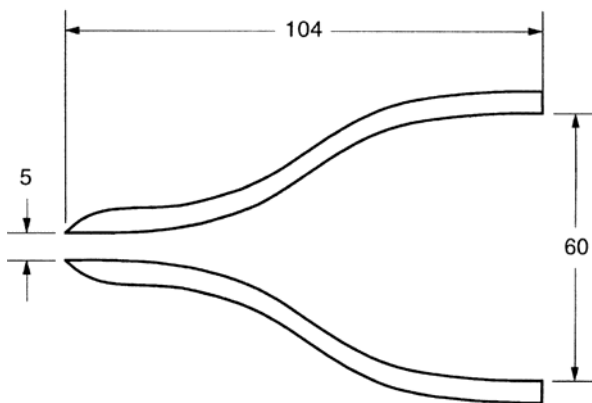


Figure 2: Nozzle Geometry

2.4. The Boundary Conditions

2.4.1. Velocity Inlet

At nozzle exit, the two components of velocity, the turbulence intensity and hydraulic diameter, $2B$, are specified. The flow is turbulent at inlet. The transverse velocity component is zero and the normal component has a flat profile. Suitable value of velocity and turbulence intensity at the inlet are chosen so that the value of the velocity and the turbulence intensity will be the same as according to the Choi et al. [15] experimental data. The constant inlet velocity in order to obtain Re_{2B} 4740, 2960 and 1780 are 0.5m/s, 0.25m/s and 0.175m/s respectively and the inlet turbulence intensity for each cases is 1%.

2.4.2. Concave Wall

The wall is non-slipping and adiabatic. Both the velocity components are zero at the wall.

2.4.3. Nozzle Wall

This wall is also non-slipping and adiabatic.

2.4.4. Pressure Outlet

The outlet condition is chosen as a constant-pressure outlet. Its position is rather far away to ensure gradual pressure drop. The ambient pressure is atmospheric. All other flow quantities at this boundary are extrapolated from inside.

2.4.5. Grid

Non-uniform quadrilateral cells are used throughout the domain. The meshing near the wall region is very fine since near the wall, the velocity gradients. Keeping in view the difficulties with generation of optimized grid for this geometry, to bring better alignment with flow direction, and to avoid convergence difficulties, the domain is meshed as shown in Figure 2. About 113,000 cells are used.

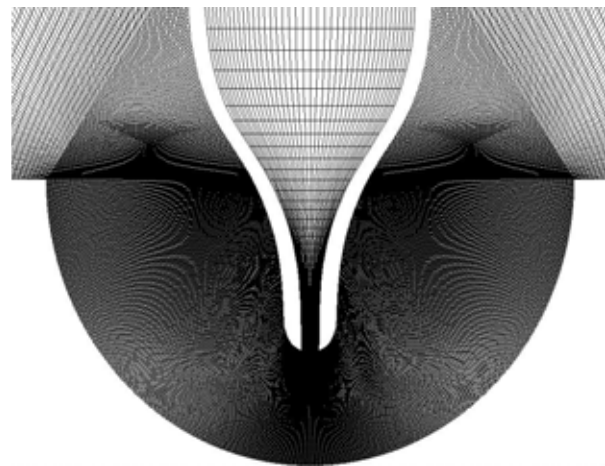


Figure 3: Meshed Domain

2.4.6. Simulation

The scaled residual for all the equations was kept at 10^{-3} . Under-relaxation factors of 0.3 and 0.7 were used for pressure and momentum respectively and default values were used for all others.

3. Results And Discussion

Impinging jet flows along the centerline have been simulated for three different Reynolds numbers equal to 1780, 2960 and 4740 as shown in Figure 4. Axial velocity, U , was non-dimensionalized by dividing it with jet velocity at the center of the nozzle exit, U_j . z is the distance along the centerline and $z = 0$ at the nozzle exit. In the figure, data from Choi et al. [15] is also included. show the

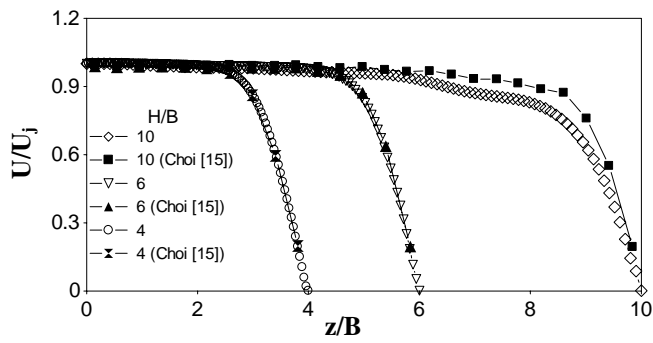
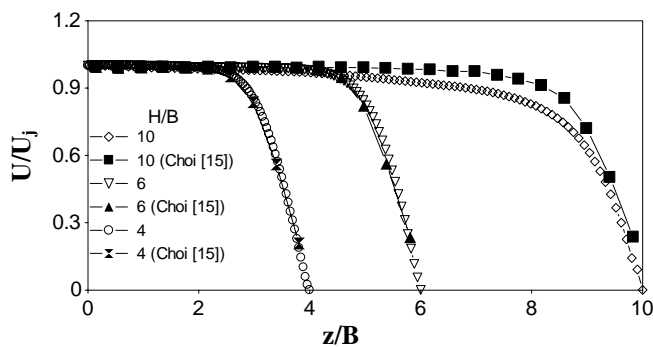
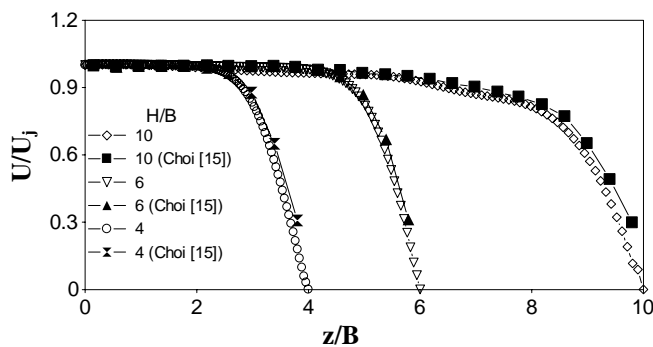
(a) $Re_{2B} = 1780$ (b) $Re_{2B} = 2960$ (c) $Re_{2B} = 4740$

Figure 4: Axial Velocity Profile Along The Centerline For Impinging Jet Flows

comparison of the velocity profile along the centerline of the impinging jet. As can be seen, the experimental data agrees very well with the model prediction. These results are obtained when using Reynolds Stress Model as the turbulence model. The results confirmed the claim of Cooper et al. [8] and Craft et al. [9] of preferring the Reynolds Stress Model for jet-impingement studies. Results shown in Figure 5 strengthen the claim. The $k-\epsilon$ and $k-\omega$ turbulence models use in the simulation are shown to give poor prediction.

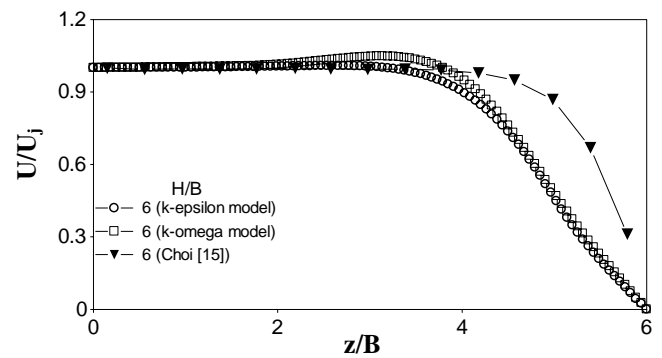


Figure 5: Performance Of $k-\epsilon$ And $k-\omega$ Model In Predicting Distribution Of Axial Velocity Profile For $H/B = 6$ At $Re_{2B} = 4740$

4. Conclusion

The main aim of the present study has been to assess the ability of computational fluid mechanics to accurately and economically predict the axial velocity distribution at the centerline of impinging jet. The following conclusion can be drawn from the numerical studies.

The $k-\epsilon$ and the $k-\omega$ model do not properly predict the velocity profile of an impinging jet. The simulations show that the Reynolds Stress Model gives good prediction of velocity distribution for jet-inlet Reynolds number of 4740, 2960 and 1780.

References

- [1] Gardon, R. and Akfirat, J.C., 1966, Heat Transfer characteristics of impinging two-dimensional air jets, *Journal of Heat Transfer*, V1.88, pp. 101-108.
- [2] Nakatogawa, T. and Nishiwaki, N. 1970, Heat transfer of round turbulent jet impinging normally on flat plate, *4th Int. Heat Transfer Conference*, pp. 1-11.
- [3] Russell, P.J. and Hatton, A. P. Turbulent flow characteristics of an impinging jet, 1972, *Proc. Instn. Mech. Engrs.*, Vol.186, pp. 635-644.
- [4] Sparrow, E. M. and Wong, T. C., Impingement transfer coefficients due to initially laminar slot jets, *Int. J. Heat Mass Transfer*, Vol. 18, pp. 597-605.
- [5] Martin, H., Heat and mass transfer between impinging gas jets and solid surface, *Advances in Heat Transfer*, 1977, Vol. 13, pp.1-60.
- [6] Looney, M. K. and Walsh, J. J., 1984, Mean flow and turbulent characteristics of free and impinging jet flows, *J. Fluid Mech.*, Vol.147, pp. 397-429.
- [7] Polat, S., Huang, B., Mujumdar, A. S. and Douglas, W. J. M., 1989, Numerical flow and heat transfer under impinging jets: a review, *Advances in Heat Transfer*, chapter 4, pp.157-197.
- [8] Cooper, D., Jackson, D. C., Launder, B. E. and Liao, G.

- X., 1993, Impinging jet studies for turbulence model assessment-I; Flow- field experiments, *Int. J. Heat Mass Transfer*, Vol.36, pp. 2675-2684.
- [9] Craft, T. J., Graham, L. J. W. and Launder, B. E., 1993, Impinging jet studies for turbulence model assessment-II. An examination of the performance of four turbulence models, *Int. J. Heat Mass Transfer*, Vol. 36, pp. 2685-2697.
- [10] Huang, L and El-Genk, M. S., Heat transfer of an impinging jet on a flat surface, *Int. J. Heat Mass Transfer*, Vol.36, pp. 1915-1923.
- [11] Chen, Q. and Modi, V., 1999, Mass transfer in Turbulent impinging slot jets, *Int. J. Heat Mass Transfer*, Vol.42, pp. 873-887.
- [12] Lee, D., Greif, R., Lee, S. J. and Le, J. H., 1995, Heat transfer from a flat plate to a fully developed axisymmetric impinging jet, *Trans. ASME Journal of Heat Transfer*, Vol.117, 772-776.
- [13] Chung, Y. S., Lee, D. H. and Lee, J. S., 1999, Heat transfer characteristics of an axisymmetric jet impinging on a rib-roughened convex surface, *Int. J. Heat Mass Transfer*, Vol.42, pp. 2101-2110.
- [14] Yang, G., Choi, M. and Lee, J. S., 1999, An experimental study of slot jet impingement cooling on concave surface: effects of nozzle configuration and curvature, *Int. J. Heat Mass Transfer*, Vol.42, pp. 2199-2209.
- [15] Choi, M., Yoo, H. S., Yang, G., Lee, J. S and Sohn, D. K., 2000, Measurements of impinging jet flow and heat transfer on a semi-circular concave surface, *Int J. Heat Mass Transfer*, Vol.43, pp. 1811-1822.

Particle Swarm Optimization Neural Network Based Modelling of Vehicle Suspension System

Gigih Priyandoko Musa Mailah Hishamuddin Jamaluddin

Department of Applied Mechanics,
Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor Bahru, Malaysia
e-mail: gigihp@gmail.com musa@fkm.utm.my hishamj@fkm.utm.my

Abstract

This paper presents a novel method for the non-parametric modelling of a passive quarter car vehicle suspension system using Neural Network (NN) that incorporates a Particle Swarm Optimization (PSO) method. The design and structure of the scheme based on this approach are described and discussed with the input and output data for a number of selected physical parameters acquired from a running quarter car test rig. Simulation results from the study show that the NN with PSO as a learning method performs better than the Backpropagation (BP) learning algorithm in executing the proposed task.

Keywords:

Vehicle suspension, modelling, neural network, back-propagation, particle swarm optimization.

1. Introduction

Modelling of a vehicle has been studied for a long time in order to design a high-performance vehicle controller. In the past, most vehicle models have been constructed analytically in the form of dynamic equations. A conventional vehicle model, however, suffers from structural complexity, long development time, and the difficulty of modelling the highly nonlinear terms and the measurement noise. In order to control a real vehicle beyond mere simulation, it is indispensable to model the vehicle quickly, easily, and accurately [1]. In paper [2], the authors develop a comprehensive nonlinear model of a vehicle suspension system. This model is derived using the standard kinematics and kinetics and takes into consideration a number of factors that are neglected in most existing models. Typical control strategies rely on linear and time-invariant models. Buckner *et al.* [3] use neural network that continually learns and estimates the nonlinear parameter variations of a quarter-car suspension model. This estimation algorithm becomes the foundation for an Intelligent Feedback Linearization controller for active vehicle suspensions. Billings *et al.* [4] demonstrated that NN could be used successfully for the identification and control of non-linear dynamical systems. The most advantageous and distinguishing feature of NN is their ability to learn. The network in the adaptive mode abstracts and generalizes the function character in the process of learning from training patterns. The learning algorithm is an optimization method capable of finding weight coefficients and thresholds (learning rates) for a given neural network and a training set. The learning algorithm that is used most frequently is the backpropagation (BP) method. Although BP training has proved to be efficient in many applications, its convergence tends to be slow, and yields to suboptimal solutions [4-5]. To counter this problem, a particle swarm optimization (PSO) technique is proposed and incorporated into the NN system. Generally, the PSO is characterized as a simple heuristic of a

well balanced mechanism with flexibility to enhance and adapt to both global and local exploration abilities. It is a stochastic search technique with reduced memory requirement, computationally effective and easier to implement [6-8].

The paper is organized as follows: Section II describes the dynamics of the vehicle suspension system while section III presents the structure of the NNs using BP learning algorithm. Section IV describes the Particle Swarm Optimization technique. The results of the simulation study are discussed and presented in section V. Finally, the paper is concluded in section VI.

2. Quarter Car Vehicle Suspension

A standard assumption in the design and analysis of controllers for vehicle suspension system is that the vertical vehicle dynamics can be modelled using four independent quarter car suspension models. Quarter car models are simple but yet capture many important characteristics of the full model. Figure 1 depicts a typical two degree-of-freedom quarter car model of a passive suspension system, in which the single wheel and axle are connected to the car body through a passive spring-damper combination, while the tyre is modelled as a simple spring. Figure 1 shows the quarter car test rig. The parameters shown in Figure 2 are defined as follows:

- m_s : sprung mass
- m_u : unsprung mass
- b_s : damping coefficient
- k_s : spring stiffness coefficient
- k_t : tyre stiffness coefficient
- z_s : displacement of the car body (sprung mass)
- z_u : displacement of wheel (unsprung mass)
- z_r : displacement of road
- $z_s - z_u$: suspension deflection

- $z_u - z_r$: tyre deflection
 \dot{z}_s : velocity of sprung mass
 \dot{z}_u : velocity of unsprung mass
 \ddot{z}_s : acceleration of sprung mass
 \ddot{z}_u : acceleration of unsprung mass

3. Neural Network

NN is basically a model structure and contains an algorithm for fitting the model to some given data. The network approach to modelling a plant uses a generic nonlinearity and allows all the parameters to be adjusted. In this way it can deal with a wide range of nonlinearities. Learning is the procedure of training the NN to represent the dynamics of a plant. The NN is placed in parallel with the plant and the error between the output of the system and the network output, the prediction error, is used as the training signal. NN have a potential for intelligent control systems because they can learn and adapt, approximate nonlinear functions, suited for parallel and distributed processing, and they naturally model multivariable systems.

The advantageous and distinguishing feature of NN is their ability to learn. The learning algorithm is an optimization method capable of finding weight coefficients and learning rate for a given NN and a training set.

This algorithm is based on minimizing the error of the NN output compared to the required output. The required function is specified by the training set. The error of network E relative to the training set is defined as the sum of the partial errors of network E_k relative to the individual training patterns and depends on network configuration w [9-10]:

$$e = \sum_{k=1}^p e_k = \frac{1}{2} \sum_{k=1}^p (O_k - d_k)^2 \quad (1)$$

where p is number of available patterns, e_k is partial network error, O_k is output of neural networks, d_k is teach data or desired output.

Updating the weights of each layer using BP method in time $t > 0$ is calculated in order to minimize error as follows [9-10]:

$$w_{jk}(t) = w_{jk}(t-1) + \Delta w_{jk}(t-1)$$

$$\Delta w_{jk}(t-1) = \alpha \frac{\partial E(t-1)}{\partial w_{jk}} + \beta (w_{jk}(t-1) - w_{jk}(t-2)) \quad (2)$$

where $0 < \alpha < 1$ is the learning rate, β is the momentum. The speed of training is dependent on the set constant α . If a low value is set, the network weights react very slowly. On the contrary, a high value is set the algorithm fails. Therefore, the parameter α is set experimentally.

The structure of the multilayer NN used in the study consists of the input, output and hidden layers. The input layer has three inputs represented by the road profile, sprung mass acceleration and suspension deflection. The output layer has two outputs, namely the sprung mass acceleration and suspension deflection. Every output neuron uses a linear

activation function. Hidden layer have three neurons and each neuron uses sigmoid bipolar activation. Figure 3 shows the architecture of the proposed NN model.

4. Particle Swarm Optimization

The PSO idea was originally introduced by Kennedy and Eberhart [6] in 1995 as a technique through individual improvement plus population cooperation and competition, which is based on the simulation of simplified social model, such as bird flocking, fish schooling and the swarm theory. Nowadays, PSO has gained much attention and wide applications in various fields [11-12].

The basic PSO algorithm consists of three steps, namely, generating particles' positions and velocities, velocity update, and position update. Here, a particle refers to a point in the design space that changes its position from one move (iteration) to another based on velocity updates. First, the positions, x_k^i , and velocities, v_k^i , of the initial swarm of particles are randomly generated using upper and lower bounds on the design variables values, x_{min} and x_{max} , as expressed in Equations (3) and (4). The positions and velocities are given in a vector format with the superscript and subscript denoting the i^{th} particle at time k . In Equations (3) and (4), $rand$ is a uniformly distributed random variable that can take any value between 0 and 1.

$$x_0^i = x_{min} + rand(x_{max} - x_{min}) \quad (3)$$

$$v_0^i = \frac{position}{time} = \frac{x_{min} + rand(x_{max} - x_{min})}{\Delta t} \quad (4)$$

The second step is to update the velocities of all particles at time $k+1$ using the particles objective or fitness values which are functions of the particles current positions in the design space at time k . The fitness function value of a particle determines which particle has the best global value in the current swarm, p_k^g , and also determines the best position p_{best}^i of each particle over time, p^i , i.e. in current and all previous group moves g_{best} . The velocity update formula uses these two pieces of information for each particle in the swarm along with the effect of current motion, v_k^i , to provide a search direction, v_{k+1}^i , for the next iteration. The velocity update formula includes some random parameters, represented by the uniformly distributed variables, $rand$, to ensure good coverage of the design space and avoid entrapment in local optima. The three values that effect the new search direction, namely, current motion, particle own memory, and swarm influence, are incorporated via a summation approach as expressed in Equation (5) with three weight factors, namely, inertia factor, w , self confidence factor, c_1 , and swarm confidence factor, c_2 .

$$v_{k+1}^i = wv_k^i + c_1 rand \frac{(p^i - x_k^i)}{\Delta t} + c_2 rand \frac{(p_k^g - x_k^i)}{\Delta t} \quad (5)$$

The position of each particle is updated using its velocity vector given by

$$x_{k+1}^i = x_k^i + v_{k+1}^i \Delta t \quad (6)$$

5. Results and Discussion

The architecture of the suspension dynamics identification using NN technique is illustrated in Figure 4. Data for training of the NN were extracted from the physical experimental rig with sprung mass weighing approximately 150 kg, unsprung mass 35 kg and tyre pressure equals to 20 psi. Sprung mass (vertical body) acceleration, suspension deflection and tyre deflection were considered as the output variables. The input variable comes in the form of the road profile. Physical sensors were incorporated to obtain the required input/output signals for identification purpose and they were connected to a PC-based data acquisition system. Accelerometer was installed at the sprung mass of the vehicle suspension to measure body acceleration, a linear variable differential transformers (LVDT) was placed in between the sprung mass and unsprung mass to measure suspension deflection and another LVDT is attached between the unsprung mass and input disturbance (road profile) to measure the vertical displacement of road profile. Two types of road profiles were used in the form of approximate sinusoidal and pulse waves as depicted in Figures 4 and 5 respectively. Both the road profiles were generated by a specially designed pneumatic system controlled by a programmable logic controller (PLC).

The training algorithm for the proposed NN in this paper is BP with PSO. It is usual that the NN parameters related to the weights, biases and learning rates (thresholds) of the BP algorithm are randomly initialized. The pseudo code of the training procedure is as follows:

```

Begin PSO
For each particle
    Initialize particle (v0 and p0)
End
Do For each particle
    Calculate fitness value
    If fitness better than pbest update pbest
End
Determine gbest from all particles
For each particle
    Update velocity to formula (5)
    Update position to formula (6)
End
While maximum iterations or minimum error criteria is not attained

Begin Neural Network
Initialize weights (wi) and learning rate θ
Do
    Input xi(t) with desired output d(t)
    Calculate error to formula (1)
    Adapt weights to formula (2)
While not done
  
```

Moreover, the parameters of the NN were determined by using PSO method in an off-line manner after a number of trial runs. Subsequently, the setting parameters of the PSO

algorithm were obtained as follows: number of particles = 50, dimensions = 20, $c_1 = 1.25$, $c_2 = 1.25$ and $w = 0.35$. Results obtained from the simulation were shown in Figures 6 and 7 related to the identified sprung mass acceleration and suspension deflection respectively for a sinusoidal wave road profile. Figures 8 to 9 are the identified sprung mass acceleration and suspension deflection respectively for a given pulse wave road profile. In all the figures, the parameters to be identified are the actual acceleration and suspension deflection curves of the quarter car represented by the blue solid lines. The black dotted lines show the identified parameters using NN with BP algorithm, while the red solid lines are those identified by the NN with PSO scheme. It is very obvious that the latter scheme performs better than the NN scheme without PSO learning algorithm since they almost replicate the actual responses.

6. Conclusions

An alternative approach to modelling a passive vehicle suspension system using NN with PSO training method has been presented and successfully applied. By using the experimental quarter car test rig data, the nonlinear characteristics of the vehicle suspension system can be captured without having to resort to its dynamic model (equations of motion). The NN model responses and the actual test rig outputs are almost identical which implies that the NN model has captured the real vehicle dynamic characteristics. Further rigorous investigation should be carried out to evaluate the proposed model performance compared with other methods. The results of this study can also be used as a basis to design more complex suspension control system involving intelligent method.

Acknowledgements

The authors would like to thank the *Malaysian Ministry of Science and Technology and the Environment (MOSTE)* and *Universiti Teknologi Malaysia (UTM)* for their continuous support in the research work. This research was fully supported by an IRPA grant (No. 03-02-06-0123EA001).

References

- [1] Yim, Y.U., Oh, S.Y., 2004, Modelling of Vehicle Dynamics From Real Vehicle Measurements Using a Neural Network With Two-Stage Hybrid Learning for Accurate Long-Term Prediction, *IEEE Trans. on Vehicular Technology*, Vol. 53, No. 4, July.
- [2] Joo, D.S., Al-Holou, N., Weaver, J.M., Lahdhiri, T., Al-Abbas, F., 2000, Nonlinear modelling of vehicle suspension system, *Proc. of the 2000 American Control Conference*, Vol. 1, Issue 6, 28-30 June.
- [3] Buckner, G.D., Schuetze, K.T., Beno, J.H., 2000, Active Vehicle Suspension Control Using Intelligent Feedback Linearization, *Proc. of the American Control Conference*, Chicago, Illinois June.

- [4] Billings, S.A., Jamaluddin, H.B., Chen, S., 1992, Properties of neural network with applications to modelling non-linear dynamic systems, *Int. Journal of Control*, vol. 55, no. 1, 193-224.
- [5] Narendra, K.S., 1996, Neural networks for control: theory and practice, *Proc. IEEE* 84, 1385-406.
- [6] Kennedy, J., Eberhart, R.C., 1995, Particle Swarm Optimisation, *Proc. IEEE International Conference on Neural Networks*, IV, 1942-1948.
- [7] Parsopoulos, K.E., Vrahatis, M.N., 2002, Recent approaches to global optimization problems through Particle Swarm Optimization, *Natural Computing*, 1, 235-306.
- [8] Van den Bergh, F., 1999, Particle Swarm weight initialization in multi-layer perceptron artificial neural networks, *ICAI*, Durban, South Africa.
- [9] Wilamowski, B.M., 2003, Neural network architectures and learning, *IEEE International Conference on Industrial Technology*, Vol. 1, 10-12 December.
- [10] Wilamowski, B.M., Chen, Y., Malinowski, A., 1999, Efficient algorithm for training neural networks with one hidden layer, *Proc. IJCNN*, vol. 3, 1725-1728.
- [11] Lu, W.Z., Fan, H.Y., Leung, A.Y. T., Wong, J.C.K., 2002, Analysis of pollutant levels in central Hong Kong applying neural network method with particle swarm optimization, *Environmental Monitoring and Assessment*, 79, 217-230.
- [12] Guo, Q.J., Yu, H.B., Xu, A.D., 2006, A hybrid PSO-GD based intelligent method for machine diagnosis, *Digital Signal Processing*, vol. 5, p. 1-17.
- [13] Hereford, J.M., Kuyucu, T., 2005, Robust neural networks using motes, *Proceedings of the 2005 NASA/DoD Conference of Evolution Hardware (EH'05)*.
- [14] Jha, R., Rower, J., 2002, Experimental investigation of active vibration control using neural networks and piezoelectric actuators, *Smart Materials and Structures*, 11, 115-121.
- [15] Wagner, J., Liu, X., 2000, Nonlinear modeling and control of automotive vibration isolation systems, *Proceedings of the American Control Conference 2000*, Volume 1, Issue 6, 28-30 June 2000 Page: 564 - 568 vol.1.
- [16] Tan, Y., Saif, M., 1997, Nonlinear Dynamic Modelling of Automotive Engines Using Neural Networks, *Proceedings of the 1997 IEEE International Conference on Control Applications*, Hanford, October 5-7.

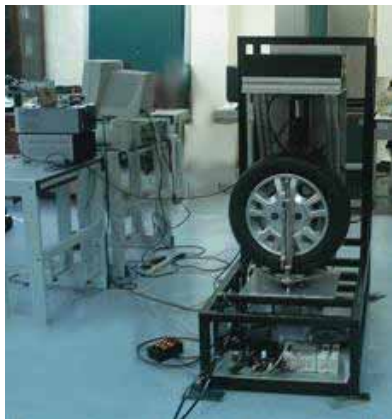


Figure 1. Quarter car test rig

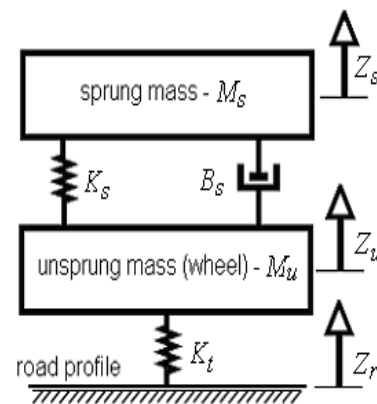


Figure 2. Quarter Car Vehicle Suspension

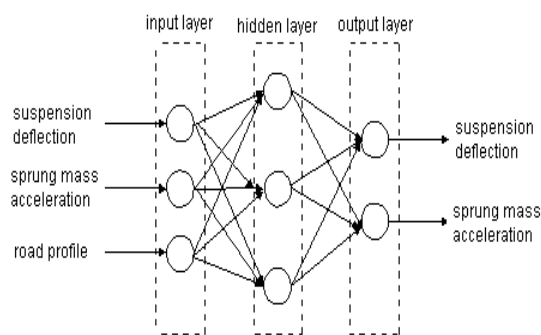


Figure 3. The architecture of the proposed NN model.

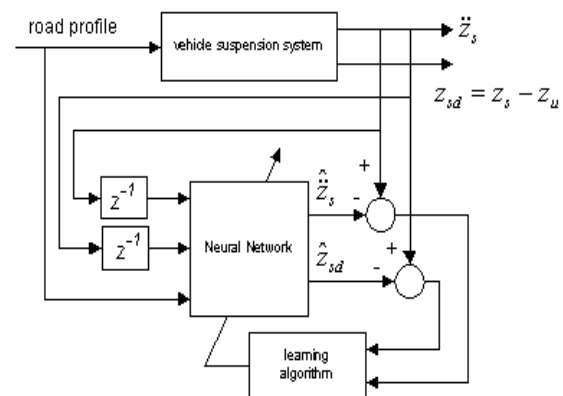


Figure 4. The architecture of the suspension dynamics modelling.

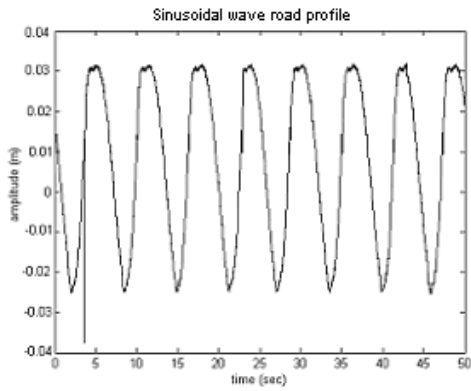


Figure 4. Sinusoidal wave road profile

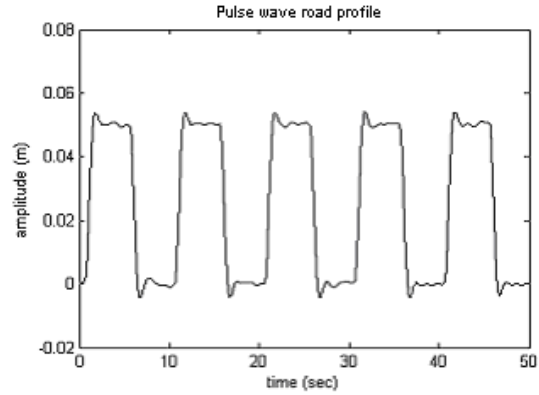


Figure 5. Pulse wave road profile

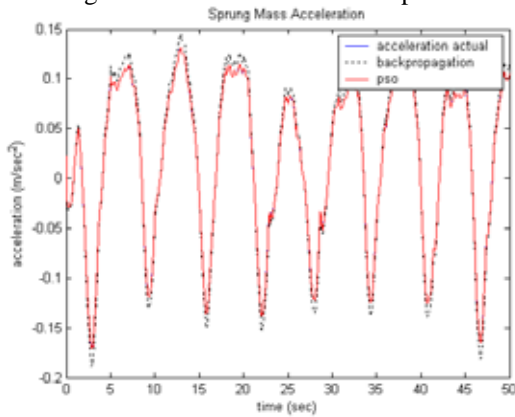


Figure 6. Sprung mass acceleration of sinusoidal wave road profile

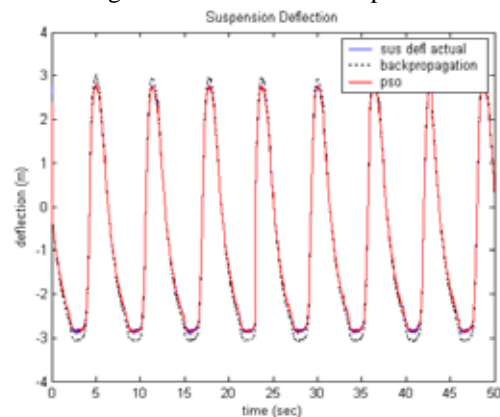


Figure 7. Suspension deflection of sinusoidal wave road profile

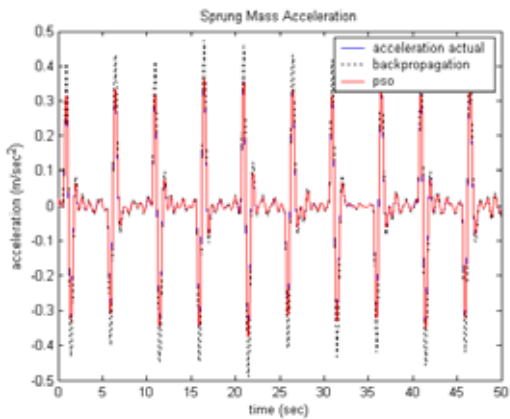


Figure 8. Sprung mass acceleration of pulse wave road profile

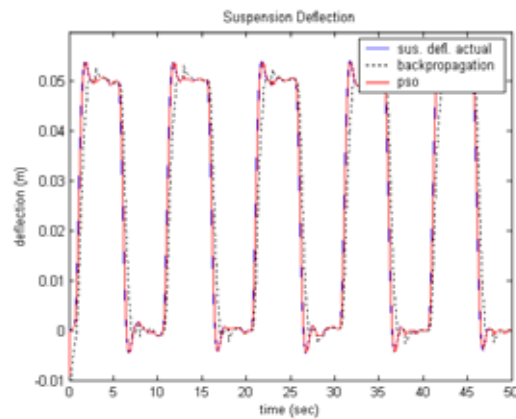


Figure 9. Suspension deflection of pulse wave road profile

Preliminary Study of Active Noise Control Method for Jet Engine Noise Using Secondary Multiple Source

Muhammad Kusni, Rudyansyah

Study Program of Aerospace Engineering,
Faculty of Industrial Technology
Bandung Institute of Technology, Bandung, Indonesia
Tel: +62-22-2504529, E-mail: kusni@ae.itb.ac.id, syahrudy@yahoo.com

Abstract

Jet engine active noise reduction concept has often been done by scientist using different kind of methods. The result gained from that method shown a good noise reduction level. One of the methods used was secondary multipole source. Active noise reduction method uses secondary multipole source, which is based on Taylor's expansion series principal to state that noise source produced by jet engine as the secondary multipole source. Secondary multipole source is a combination of quadrupole, octopole order, and so on. Noise area in jet engine can be reduced by revising the secondary multipole source phase. Noise field in jet engine can be illustrated as quadrupole field pattern. This noise will be reduced using active noise control. Jet engine noise is illustrated as quadrupole primary source, while for the anti-noise source secondary multipole source is used. The numeric simulation for this method was done by limiting the Taylor's series order. This was done because it was impossible to use secondary multipole source with infinite order for the implementation. The result of the simulation shows that sound reduction can reach 30 dB for 205 Hz frequency. This simulation was done in free field with single frequency.

Keywords:

noise control, jet engine noise, quadrupole source

1. Introduction

Nowadays, noise control matter has become a crucial problem for the society. Other than consider as environmental pollution, this noise matter could also bring physical and biological effect to human such as affecting concentration, verbal communication interruption, and it could also decreases hearing ability, temporally or permanently. Therefore, noise control aspect has become important to be solved, so the problems because of the noise are handling with care.

There are several ways to do noise control, such as reducing noise actively or passively. Active noise reduction method is based on the superposition destructive wave principal. On the other hand, passive noise method is based on design transformation in sound source principal or by adding absorber or barrier on a sound source.

Based on international regulation that controls the airplane's exterior noise level FAR (Federal Aviation Regulation) part 36, the principal of active noise control become an important thing to be done by flight companies. Therefore, this study will talk about active noise control to reduce noise source caused by jet engine modeled in form of quadrupole field pattern.

In active noise reduction method, the noise is contrasted with anti-noise. Anti-noise source has similar

amplitude and frequency with the noise, but its phase is different about 180°. So, both of the sources will be weaken each other. By arranging the characteristic of the anti-noise source, it will be able to reduce noise in a certain period of frequency and field according to the needs. However, this method has problems on its testing, such as the cost in computing is highly expensive, the lack of effectiveness in reducing noise in high frequency, and the location of secondary source which has to be $\frac{1}{2}$ the length of the highest frequency that needs to be reduced from the primary source just to have an efficient reduction. Nevertheless, this study will show that those problems can be solved using active noise reduction which uses secondary multipole source.

2. Approach and Methods

2.1 The Concept of Secondary Multipole

The concept about multipole was created by Kempton in 1976. According to Kempton, monopole sound field can be produced in exact by putting multipole source with infinite order. He showed that the field that can be enclosed with exact is in the area outside a sphere that rotates on the multipole source, and that still covers the primary monopole source, and that still covers the primary monopole source. Because the noise source that will be reduced is a jet noise

that has quadrupole pattern (figure 1), the secondary source that will be used as anti-noise is a secondary multipole source.

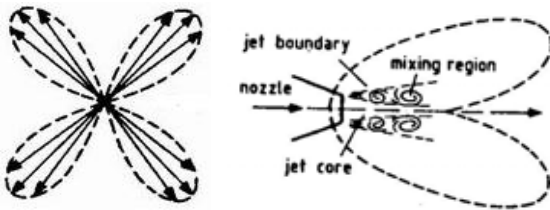


Figure 1. Quadrupole and Jet Noise Pattern

The analysis done to gain the mathematics model stating the sound pressure because of the noise source or anti-noise source uses the assumption that the position used as the viewing point is on the far field, so the primary quadrupole and secondary multipole source can be seen as the source in form of dots. Every analytical function $g(x)$ can be called in this Taylor's form:

$$g(x+h) = \sum_{n=0}^{\infty} \frac{h^n}{n!} \frac{\partial^n}{\partial x^n} g(x) \tag{1}$$

Whereas h denote the incremental rising of the variable x and apply on the area $|x| > |h|$

The secondary multipole source is produce by separating the equation, which states the sound pressure because of the noise into Taylor's series on around the noise. Jet engine noise is put in the position $-h$, then, the noise is restated by doing the approach using Taylor's series on the central coordinate, or in other word, the secondary multipole source is placed on the central coordinate (figure 2).

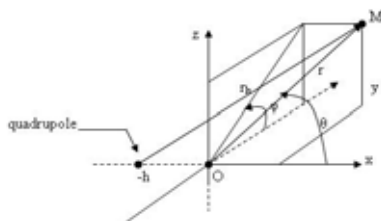


Figure 2. Geometry and Coordinate System

After having the equation named the sound pressure in point M for every source, then the equation that tells the strength of every component in the secondary multipole source would have to be found. By knowing the strength, then an anti-noise that has certain source power hoped to reduce the jet's noise can be made. To gain the secondary multipole source power, two approaches are used. The first approach is comparing secondary multipole source to the jet's noise in Taylor's series. This approach is called direct approach. The second approach is doing optimization of the source's power. The way is by searching the total minimum sound energy (combination of noise source and anti-noise).

Simulation in order to gain the sound reduction is using the equation of sound pressure level. That is

comparing the total sound pressure after being reduced to the noise before the reduction.

2.2. Secondary Multipole Source

Sound pressure in point M caused by jet's noise source is stated with the equation:

$$p_{pq}(x, y, z) = \frac{i\omega\rho}{4\pi} Q_{pq} \frac{\partial^2}{\partial x^2} \frac{e^{ikr_h}}{r} \tag{2}$$

If the equation is restated by separating it in Taylor's series, then an equation that states the secondary multipole source will be gained. Next, if the secondary source is operated to a contrasting phase from the noise, a sound weakening will occur. The explanation of the equation (2) in Taylor's series is like this:

$$p_{pq}(x, y, z) = \frac{i\omega\rho}{4\pi} \sum_{n=0}^{\infty} Q_{pq} \frac{h^n}{n!} \frac{\partial^{n+2}}{\partial x^{n+2}} \frac{e^{ikr}}{r} \tag{3}$$

The used order are qudrupole and octopole, that is for rate of $n=0$ and $n=1$. The sound pressures for each component in secondary multipole source are:

$$p_{sq}^f(r, \theta) = \frac{i\omega\rho}{4\pi} k^2 \cos^2 \theta Q_{sq} \frac{e^{ikr}}{r} \tag{4}$$

$$p_{so}^f(r, \theta) = \frac{\omega\rho}{4\pi} k^3 \cos^3 \theta Q_{so} \frac{e^{ikr}}{r} \tag{5}$$

2.3. The Power of Secondary Multipole Source Using Direct Approach

From the equation (3) it can be observed that for the rate $n=0$, the secondary quadrupole source would have the same as primary quadrupole power source which is $Q_{s0}=Q_{pq}$, and for $n=1$ the power of secondary octopole is $Q_{s1}=h Q_{pq}$. If the equation is stated in general form will results

$Q_{sn} = \frac{h^n}{n!} Q_{pq}$. The method to find the secondary multipole power source is by directly comparing the separation of Taylor's series, and then called as direct approach.

2.4. Secondary Multipole Power Source Using Optimized Approach

Secondary multipole power gained using direct approach was not representative enough, because it represents one series of component only. To gain more accurate result, the other component series would have to be included in order to find the rate of multipole power source. The total sound pressure and the radial velocity for every order in multipole source in sequence are:

$$P_{t0} = \frac{i\omega\rho}{4\pi} \frac{e^{ikr}}{r} k^2 \cos^2 \theta [Q_{pq} e^{ikh \cos \theta} + Q_{sq}] \tag{6}$$

$$P_{t1} = \frac{i\omega\rho}{4\pi} \frac{e^{ikr}}{r} k^2 \cos^2 \theta [Q_{pq} e^{ikh \cos \theta} + Q_{sq} ik \cos \theta Q_{so}] \tag{7}$$

$$V_{r,0} = \frac{ik e^{ikr}}{4\pi r} k^2 \cos^2 \theta [Q_{pq} e^{ikh \cos \theta} + Q_{sq}] \quad (8)$$

$$V_{r,1} = \frac{ik e^{ikr}}{4\pi r} k^2 \cos^2 \theta [Q_{pq} e^{ikh \cos \theta} + Q_{sq} ik \cos \theta Q_{so}] \quad (9)$$

The radial sound intensity would then be stated using the equation:

$$I = \frac{1}{2} Re(p^* v) \quad (10)$$

Whereas * state the conjugate.

Sound energy is gained by integrating equation (10) along the sphere surface, which is:

$$W = 2\pi r^2 \int_0^\pi I_r \sin \theta d\theta \quad (11)$$

After the sound energy is gained, the equation resulted is dropped to the secondary power source searched in order to have the minimal sound energy. Those equations will result the secondary multipole power source. The equation that stated the secondary multipole power source for quadrupole order until octopole are as follow:

- For quadrupole order, the multipole component consists of quadrupole only.

$$Q_{sq} = 20 \frac{\sin(kh)}{4(kh)} + \frac{\cos(kh)}{(kh)^2} 3 \frac{\sin(kh)}{(kh)^3} 6 \frac{\cos(kh)}{(kh)^4} + 6 \frac{\sin(kh)}{(kh)^5} Q_{pq} \quad (12)$$

- For octopole order, the multipole component consists of quadrupole and octopole.

$$Q_{sq} = 20 \frac{\sin(kh)}{4(kh)} + \frac{\cos(kh)}{(kh)^2} 3 \frac{\sin(kh)}{(kh)^3} 6 \frac{\cos(kh)}{(kh)^4} + 6 \frac{\sin(kh)}{(kh)^5} Q_{pq} \quad (13)$$

$$Q_{so} = \frac{7}{2k} 2 \frac{\cos(kh)}{(kh)} + 10 \frac{\sin(kh)}{(kh)^2} + 40 \frac{\cos(kh)}{(kh)^3} 120 \frac{\sin(kh)}{(kh)^4} 240 \frac{\cos(kh)}{(kh)^5} + 240 \frac{\sin(kh)}{(kh)^6} Q_{pq} \quad (14)$$

3. Result and Discussion

3.1 Sound Pressure Reduction

The result of sound pressure reduction is stated using total pressure ratio after reduced to the primary source pressure before reduction.

$$L_c = 20 \log \left| \frac{P_m}{P_{pq}} \right| \quad (15)$$

Whereas $P_m = P_{pq} + P_{sn}$ states the sound pressure after using active multipole reduction. P_{pq} states the sound pressure resulted by quadrupole primary source, while P_{sn} states the sound pressure resulted by multipole secondary

source that is cut on the order if it is operated reversed phase to quadrupole primary source. So L_c will worth negative if the level of primary sound pressure source is reduced by secondary source.

Sound pressure reduction for direct approach gives the equation:

$$L_c^f = 20 \log \left| \frac{e^{ikh \cos \theta} \sum_0^\infty \frac{(ikh \cos \theta)^n}{n!}}{e^{ikh \cos \theta}} \right| \quad (16)$$

Data used for the calculation on the figure (3) are: frequency = 250 Hz, distance between primary source and secondary source = 0.34 m (1/4 the wave length), and the sound speed is 340 m/s, in this case the value of kh is 1.57.

Figure (3) shows that the level of sound reduction will be rise, if the secondary multipole sources order enlarged. For quadrupole source, sound strengthening still happens in 0° to 48° , 132° to 228° , and 312° to 360° . In octopole order, sound strengthening still happens in small period, such as 0° to 21° , 159° to 201° , 339° to 360° . In quadrupole and octopole order, the maximum reduction happens on the 90° and 270° .

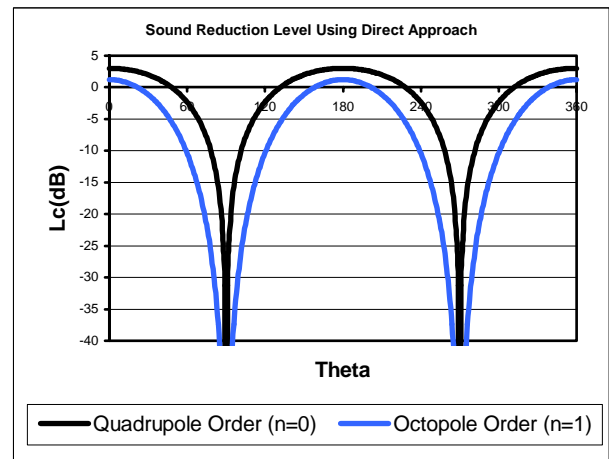


Figure 3. Sound Reduction Level (L_c) Vs θ Using Direct Approach

Figure (4) shows that the sound reduction is depend on the angle of θ . The smallest sound gained on axis x , which is the axis that connects primary and secondary source ($\theta = 0^\circ$ and $\theta = 180^\circ$). It also shows that if the distance of r is getting bigger, the sound reduction will not depends on the distance of r .

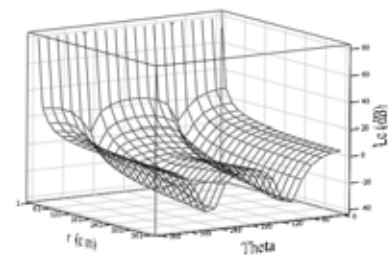


Figure 4. Sound Reduction Level as Function θ and r (Direct Approach $n=1$)

While in figure (5) appeared that on the axis x , sound reduction is still effective until frequency more than 500 Hz, as long as the value of h is big enough, as long as it is operated in low frequency.

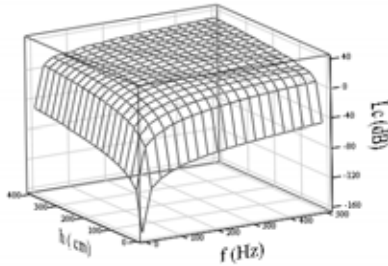


Figure 5. Sound Reduction Level as Function f and h (Direct Approach $n=1$)

Calculation and simulation procedure done for the method of source power optimization is similar with direct approach. Sound pressure reduction is gained by rationing the sound pressure after reduction to the sound pressure of quadrupole primary source.

In figure (6), it is seen that the method of source power optimization gives a better reduction result compared to direct approach. Other than it can reduce a bigger sound pressure, the reduction result is also varies.

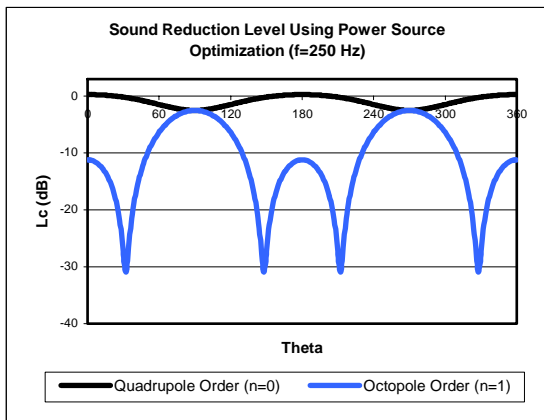


Figure 6. Sound Reduction Level (L_c) Vs θ Using the Method of Power Source Optimization

The comparison between figure (8) and (5) shows that the result of sound reduction using secondary power source optimization gives a better result if compared to what it gained by using direct approach. Using the method of power source optimization, the sound reduction could be reached until higher and farther frequency between primary and secondary source.

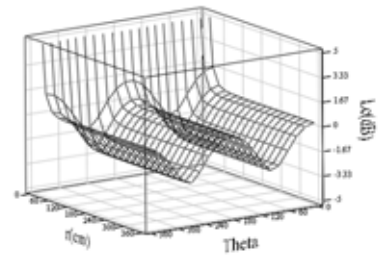


Figure 7. Sound Reduction Level as Function θ and r (Power Source Optimization Method $n=0$)

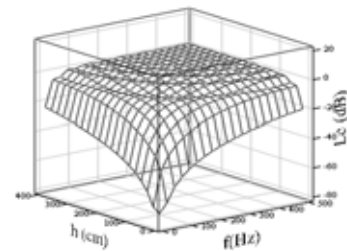


Figure 8. Sound Reduction Level as Function f and h (Power Source Optimization Method $n=1$)

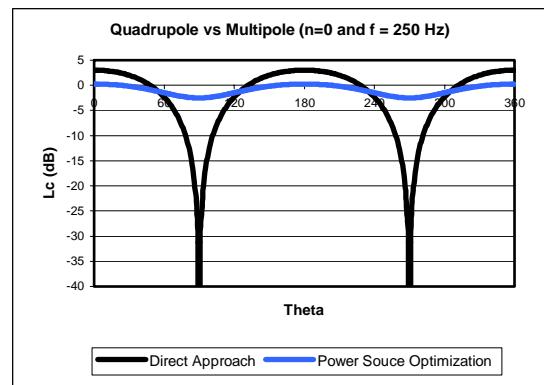


Figure 9. Comparison between Direct Approach and Power Source Optimization Method for Quadrupole Order

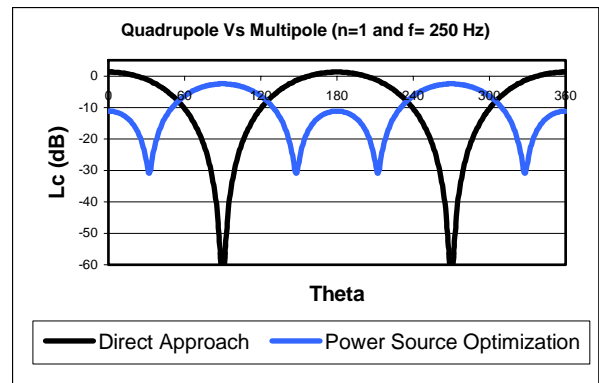


Figure 10. Comparison between Direct Approach and Power Source Optimization Method for Octopole Order

4. Conclusion

The approach method done caused the multipole secondary source used to have infinite order. This comprises the consequence that the reduction result gained is also limited. The method for power source optimization was done to get a better sound reduction result than direct approach.

By optimizing the secondary multipole power source, a relatively good sound reduction will be gained even though the order used is limited to octopole only. The comparison between those methods can be seen on figure (9) and (10). Based on that figure, it can be concluded that the sound reduction gained by optimizing power source is more varies. Other than it can be used in noise with high frequency, the distance between primary noise (jet) and secondary multipole source can be widen.

Acknowledgments

This work was supported by Study Program of Aerospace Engineering, Bandung Institute of Technology.

References

- [1] Kusni, M. 2001. Active Noise Control Method Using Secondary Multipole Source. Thesis, Dept.of Physics Engineering, Bandung Institute of Technology.
- [2] Saputra, 2004. Active Noise Control Method of Propeller Noise Source in Free Sound Field Using Secondary Multipole Source. Thesis, Dept. of Aerospace Engineering, Bandung Institute of Technology.
- [3] Purnomo, 2003. Preliminary Implementation of Active Noise Control Method Using Secondary Multipole Source. Thesis, Dept. of Aerospace Engineering, Bandung Institute of Technology.
- [4] Ruijgrok, G.J.J. eds. 1993. *Elements of Aviation Acoustics*. Delft University Press.
- [5] Golgstein, Marvin E. eds. 1976. *Aeroacoustics*. McGraw-Hill.

Optimum Number of Stages of the New Multi-Stage Symmetrical Wobble Plate Compressor

Ardiyansyah Syahrom^{1*}, Md. Nor Musa², Wan Ali Wan Mat², Ainulotfi Abdul Latif⁴

^{1,2,3,4} Faculty of Mechanical Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5534878, Fax: +607-5566159,
E-mail : ¹ardiyansyah@utm.my
²mdnor@fkm.utm.my
³wanali@fkm.utm.my
⁴lotfi@fkm.utm.my

Abstract

A new multistage symmetrical wobble plate compressor had been developed. Working pressure is from 3 bar to 206 bar and to compress natural gas for vehicle (NGV). This paper describes on how to get the optimum number of stages of this new Multi-stage symmetrical wobble plate compressor. There are several parameters involved such as load of compressor, torque, work, pressure ratio, and finally the overall size of the compressor. In this design, number of stages start from 3 (three) to 7 (seven). The optimum number of stages optimum for capacity 10 m³/hr at for pressure ratio of 2.27 the optimum number of stages is five certain pressure ratio could be obtained.

Keywords:

Symmetrical Wobble Plate Compressor, Wobble Plate, Compressor, High Pressure, NGV

1. Introduction

For allowable number of stages chosen and fixed clearance at certain pressure ratio, discharge temperature and efficiency could be calculated. If the calculated discharge temperature in one stage is too high, therefore more stages are required. During preliminary sizing, the isentropic discharge temperature could be used, but if a certain number of stages create a marginal situation, the discharge temperature should be estimated more accurately. As a first estimate, it can be assumed that equal pressure ratios are used for all stages. In practice it is often better to take a higher-pressure ratio in the low-pressure ratio stages and unload the more critical higher-pressure stages a little.

If stages number is to high, the pressures losses in the valve and piping will offset the gains from inter cooling and the efficiency would be reduced. The cost of the compressor to do a given task usually increases as the number of stages is increased because of the additional compressor cylinder, coolers and piping needed.

In this case, a symmetrical wobble plate compressor could be used to compress a natural gas. An operating conditions are suction pressure 50 psi (3 bar), discharge pressure 3000 psi (206 bar), and capacity 10Nm³/hr. There are several parameters involve in getting the optimum number of stages i. e., load of compressor, torque, work,

pressure ratio, and finally is overall size of the compressor.

In this design process the number of design stages will be start from 3 (three) until 7 (seven) and the optimum number of stages could be obtained at certain pressure ratio.

2. Pressure Ratio

If the compression ratio is increased, the final compression temperature rises and the volumetric efficiency of the compressor decrease. A high compression temperature affects the operation of the delivery valve, diminishes the lubrication properties of the oil and involves. The maximum compression ratio for the small single-stages compressor is usually 8 (eight) and for large machines are 5 (five). In the case of multi-stages compression, gas leaving the second stage is again cooled in second intercooler before being passed to the third stages and this process is repeated until the required pressure is reached. By cooling the gas between stages, the compression process approach to isothermal. Thus the final temperature is appreciably lowered and the work required for the compression is reduced.

The total work is the sum of the low-pressure and the high-pressure work.

$$-W' = \dot{m} R T_1 \frac{n}{n-1} \left[\left(\frac{P_d}{P_s} \right)^{\frac{n-1}{n}} - 1 \right] + \dot{m} R T_1 \frac{n}{n-1} \left[\left(\frac{P_d}{P_a} \right)^{\frac{n-1}{n}} - 1 \right] \quad 1$$

or

$$-W' = mRT_1 \frac{n}{n-1} \left[\left(\frac{P_a}{P_s} \right)^{\frac{n-1}{n}} + \left(\frac{P_d}{P_a} \right)^{\frac{n-1}{n}} - 2 \right] \quad 2$$

A minimum total work could be obtained by setting the first derivative coefficient of the expression in the brackets with respect to P_a of equation 1 is zero. After substituting $z = \frac{n-1}{n}$, we obtain:

$$\frac{zP_a^{z-1}}{P_1^z} - \frac{zP_d^z}{P_a^{z+1}} = 0 \quad 3$$

And therefore,

$$P_a^{2z} = P_s^z P_d^z \quad 4$$

Extracting the root leads to:

$$P_a = \sqrt{P_s P_d} \text{ or } \frac{P_a}{P_s} = \frac{P_d}{P_a} \quad 5$$

The pressure ratio in each level should be the same in order to maintain a lower total power input. So, the second equation becomes:

$$-W' = 2 \frac{n}{n-1} mRT_1 \left[\left(\frac{P_a}{P_s} \right)^{\frac{n-1}{n}} - 1 \right] \quad 6$$

If each the stage of two stages compressor subdivided into a further two stages, then have obtain the four-stage compressor. The total work of the first and second stages will be a minimum if compression ratios are the same in both stages. This also applies to the third and fourth stages. Since the compression ratio in a both stages of the original two-stage compression was equal, the compression ratios of all four stages will also be equal too. Let us now illustrate adiabatic four-stage compression on the T-s diagram (1). Assuming perfect inter cooler, and then the final compression temperature for all stages are equal. Hence the entropy changes in all stages are also equal i.e. $(s_0 - s_1) = (s_1 - s_2) = (s_2 - s_3) = (s_3 - s_4)$.

In general for the compressor with n stages will give;

$$\frac{P_1}{P_0} = \frac{P_2}{P_1} = \dots = \frac{P_n}{P_{n-1}} = \sqrt[n]{\frac{P_n}{P_0}} \quad 7$$

$$P_n = P_{n-1} \cdot \text{Pressure Ratio}$$

The equation number 8 (eight) will be use to determine the level of compressor pressure in each level.

Fig.1 and 2 show the increase of temperature in the compressor from T_s to T_a due too. If there is a cooling system and the temperature T_a will be decrease to a T_a' , then this temperature T_a' will increase become T_2 due to compression process. Generally, the temperature in delivery of compressor could be determined by using the equation as follow:

$$T_n = T_{n-1} \left(\frac{P_n}{P_{n-1}} \right)^{\frac{n-1}{n}} \quad 8$$

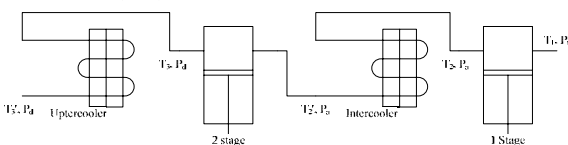


Fig.1 Plan showing inter cooling and upercooling between compressor stages

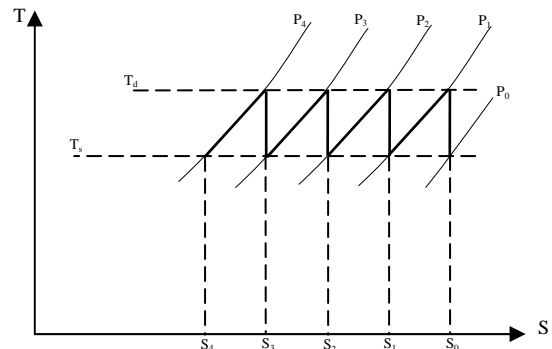


Fig.2 Adiabatic four-stage compression on the T-s diagram

3. Motion Symmetrical Wobble Plate Compressor

The governing equations for piston movement, connecting rod, wobble plate could be determined by using standard coordinate transformation method derives anti-rotation mechanism. These equations are written and used to demonstrate the effects of anti-rotation mechanism and variation in stroke length, equation no (8).

3.1 Wobble Plate Motion

Wobble plate compressors exhibits complex motion compare to crankshaft reciprocating compressors and swash plate compressor. In swash plate compressors, piston movements are sinusoidal and they only affected by the plate movement in shaft axis direction (z-axis) only. However, in wobble plate compressor all the three direction of the wobble plate movement whose are x-axis, y-axis, and z-axis will affect the movement of piston due to the usage of ball joint at connecting rod to connect between plate and piston.

Wobble plate kinematics is obtained from the movement of connecting rod. Connecting rod ball joint at wobble plate side will show the movement of plate while the connecting rod ball joint at piston side will show the movement of piston. The plate is constrained from rotating with rotor and shaft using the anti rotating mechanism. This constrain is needed to prevent connecting rod from being tangled and tied together.

The piston stroke is given by:

$$l = (z_t)_{max} - (z_t)_{min} \quad 9$$

where:

$$z_s = ((t_w) \cos\gamma + (R_w \sin\theta_n) \sin\gamma) \cos\alpha + (R_w \cos\theta_n) \sin\alpha$$

$$z_t = ((t_w) \cos\gamma + (R_w \sin\theta_n) \sin\gamma) \cos\alpha + (R_w \cos\theta_n) \sin\alpha + l_{cr} \cos\phi_n$$

From previous sets of equation, piston stroke is a function of:

$$l = f(R_w, R_p, t_w, l_{cr}, \theta_n, \tilde{\alpha}) \quad 10$$

4 Cylinders Symmetrical Wobble Plate Compressor Sizing

Once the number of stages of stages is selected, the cylinders for each stage could be obtained. Usually a selection could be made from the cylinder design. A cylinder

bore could be calculated once inlet conditions, capacity, speed and stroke are known. The available cylinders then be use to check whether, meets it the requirement. The following must be checked.

First, the pressure rating of the cylinder must be adequate for safe design at any upset conditions. The cylinder pressure rating should be higher than the relief valve setting. Second, the frame load, rod load and degrees of reversal must be within the rating for the frame components. Third, the capacity calculated with the minimum cylinder clearance allowing for all losses must meet the requirements. Fourth, the power requirement of this cylinder must not exceed the power rating per throw of the frame components.

If all this requirements fulfill, a suitable cylinder could be chosen. Additional optimization may be needed to determine the best possible ready made cylinder for this application. If none of them is suitable, then a new cylinder should be redesigned. A frame rated for a higher frame load or horsepower per throw must be selected. Two or more, smaller cylinder must be chosen to run in parallel to meet the required flow. Note that if smaller cylinders are used, the frame load and the power per throw will be reduced. Usually for smaller cylinders available in higher-pressure ratio versions, so all the requirements can usually be met by using multiple cylinders per stage.

This basic compressor sizing is now completed, but we must check with alternate design or upset conditions. Additional factor must also need such as the out-of-balance force transmitted from the compressor to the foundation, the potential for harmful torsion vibration in the shaft and driven train, optimization of the compressor layout, efficiency, and the cost involved.

4.1 Determination of Cylinders Volume

There are several factors influence in determine of the cylinders size such as radius of wobble plate, tilting angle, stroke of compressor, speed of compressor, and space for cylinders arrangement. All of these factors cannot deliberate one by one but it must be deliberated as a united. This is due to all those factors were affected by each other. For a sample, if the tilting angle is makes it big, other parameter such as the wobble plate radius also will becomes smaller.

The determination of the cylinder symmetrical wobble plate dimension was very unique. One of them is the availability of space for cylinders arrangement as describe in Fig.3.

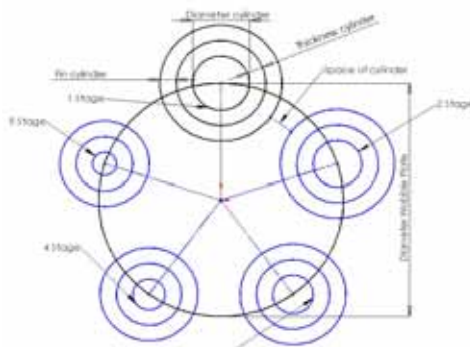


Fig.3 Cylinder configuration

Based on Fig.3, it shows that the diameter of the cylinder configuration is depending on cylinder diameter, fin diameter, and the fin thickness.

The first step in the optimization process is to determine the appropriate dimension and the configuration to fulfill $10Nm^3/hr$ capacity. Then continue to determine the optimum of the tilting angle. The same calculation as to determine the optimum number of stages, being used to determine the tilting angle. It shows that.

$$\dot{V} = D.l.speed \text{ of compressor} \quad 11$$

l is sign for the stroke of compressor and it could be calculated a by using the equation of piston stroke.

$$V_1 = \frac{\pi}{4} (D_1)^2 \quad 12$$

The next diameter could be obtained by using the following equation.

$$V_n = V_{n-1} \left(\frac{P_n}{P_{n-1}} \right)^{\gamma} \quad 13$$

A cylinder diameter of the second stage could be determined once we know the volume from equation (12). Then we have to check the availability of space to fix the cylinder.

4.2 Force Acting on the Piston

A design process of the wobble plate mechanism of the compressor could be started with the force acting on the piston. The force on piston is depend the pressure of the gas. The pressure inside the cylinder may vary and depending that appropriate an angle of rotating on of the shaft, which is started with suction process, compression and finally discharge.

For the multistage compressor, the forces in each piston have variations appropriate with the pressure and the cylinder wide. It cause of each cylinder to have different geometry and also the pressure in each stages. On the other hand, for the single stage compressor the force in each piston is the same due to the geometry and the pressure in each cylinder are the same too. For process in one time we could see that the pressure in each compressor will be different because there is a different process in each cylinder i.e. first cylinder in suction condition while the next cylinder (second cylinder) be in compression and discharge process. All of those things being the causes of the pressure differences for each cylinder. As general the force in the piston could be determined by using the equation below.

$$P = \frac{F}{A}$$

By knowing the pressure and the wide of the cylinder, we can calculated the force on the piston with equation below:

$$F = P.A \quad 14$$

Compressors for wide range of application tend to run with about the same piston speed. That is compressors with a long stroke tend to run slow than those with a short stroke. Further, short stroke compressors tend to be of lighter construction with lower allowed loads. For the best efficiency and reliability at the expense of increased cost, a piston speed at the low end of the normal rang will be used. The compressor speed and the stroke will then could be determined by the horsepower requirement. A low

horsepower compressor will require a light, low stroke, and high-speed. A higher horse power compressor will require a heavy, long strike, low speed. If possible, large compressors are normally directly couple to the prime mover of drive. Thus the speed range of available drivers may influence the selection of the compressor too.

4.3 Torques in Compressor

Torque in compressor could be determined once we know the force on each piston and the distribution of force in each angle of rotating shaft. There are several why reasons we have calculate the torque in the compressor. First, optimizing the use of energy when the compressor in running condition. Second, choose a suitable motor to run the will increase a compressor. The optimum torques the compressor performance better.

To determine the torque on the symmetrical wobble plate is totally different compare to compressors. It is describe by force diagram and torque as follow:

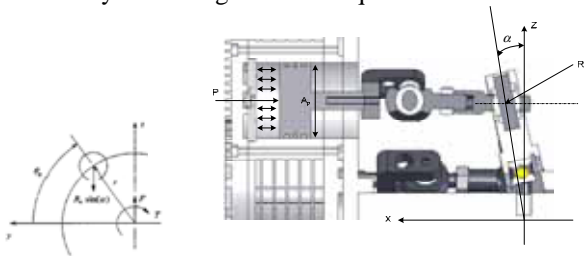


Fig.4 Force and torque diagram for loads exerted on the shaft

Fig.4 shows a sectioned view of a single piston within the cylinder block as it operates within the compressor. In this view, the reaction force between the wobble plate and the *n*th slot is shown by the symbol, *R_n*. From the left hand side of 0, it can be seen that the component of this force in the downward direction is given by:

$$F_n = R_n \sin \alpha \tag{15}$$

Where α the wobble plate angle of the compressor.

Equation 14 represents the downward force exerted on the shaft by the *n*th piston. Summing these forces for all pistons within the compressor yields the total force exerted on the shaft in the downward direction. This result is given by:

$$F = \sum_{n=1}^N R_n \sin \alpha \tag{16}$$

Where *N* is the total number of pistons within the compressor.

The torque on the shaft is generated by the downward component of the reaction force between the *n*th piston and the wobble plate, multiplied by the distance of the *n*th piston from the *z*-axis. From Fig. 5, The piston is located at a distance (from the *z*-axis) by the expression, *r cos (θ_n)*, where *r* is a piston pitch radius and θ_n is the circular position of the *n*th piston. Multiplying this distance by the right hand side of the equation 14 yields the following result for the torque exerted on the shaft by the *n*th piston, as:

$$T_n = R_n \sin \alpha r \cos \theta_n \tag{17}$$

Summing this torques for all pistons within the compressor yields the total torque exerted on the shaft as;

$$T = \sum_{n=1}^N R_n \sin \alpha r \cos \theta_n \tag{18}$$

Fig 5 to illustrate graphically the total downward force is given by Equation 15. while the total of torque is given by equation 17.

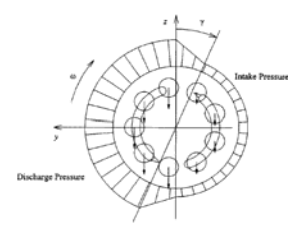


Fig.5 Piston pressure profile

The reaction between the *n*th piston and the wobble plate may be determined by summing the forces which are acting on the piston in the *x*-direction and setting equal to the piston's time rate-of-change of linear momentum. Writing this equation, and rearranging terms, yields the following result for the reaction force between the *n*th piston and the wobble plate:

$$R_n = \frac{M_p \cdot \ddot{x}_n + P_n \cdot A_p}{\cos \alpha} \tag{19}$$

where *M_p* is the mass of a single piston, \ddot{x}_n is the piston's acceleration in the *x*-direction, *P_n* is the fluid pressure within the *n*th piston chamber, and *A_p* is the pressurized area of a single piston.

The general expressions for the downward force and torque exerted on the shaft are given in Equation 15 and 17. These equations describe the instantaneous loads that are exerted on the shaft, which tends to oscillate at certain dominant frequencies depending upon the number of pistons within the compressor and the rotational speed of the shaft. If the Compressor is designed with a sufficiently large number of pistons, where the amplitude of the oscillations could be reduced and the frequency of the oscillations could be increased.

In general, for a 9-piston compressor, the amplitude of the oscillations which occur for the instantaneous downward-force are less than $\pm 10\%$ of the average force while the amplitude of the oscillations which occur for the instantaneous torque are less than $\pm 1.5\%$ of the average torque. Again, the size of these amplitudes will vary depending upon the number of pistons, which are used in the compressor design. If we assumed the oscillatory effects of the instantaneous force and torque are small compared to their average effects, then the oscillatory effects may be neglected and the average force and torque may be used to consider the loading of the shaft.

Bu using Equation 15 and 17, the average quantities of force and torque may be computed using the integral-averaging technique. This technique yields the following general forms:

$$\bar{F} = \frac{N}{2\pi} \int_0^{2\pi} R_n \cdot \sin\alpha \cdot d\theta_n \tag{20}$$

$$\bar{T} = \frac{N}{2\pi} \int_0^{2\pi} R_n \cdot \sin\alpha \cdot r \cos\theta_n \cdot d\theta_n \tag{21}$$

By discontinuous integration of the pressure terms equation 19, yields the following results for the average force and torque exerted on the shaft as;

$$\bar{F} = \frac{NA_p (P_d + P_i) \tan(\alpha)}{2} \tag{22}$$

$$\bar{T} = \frac{NA_p (P_d + P_i) r \tan(\alpha)}{\pi} \tag{23}$$

5 Result and Discussion

The stages had been analyzed from 3 (three) until 7 (seven) and that compressor use cooling system in each inter and after stages. Then, it just continues to determine the source for better number of the stages that is better. The pressure ratio for each stage could be determined as describes below.

Pressure ratio and pressure in each stage given in Table 1;

Table 1. Pressure ratio and pressure each stages.

No	No of Stage	Press Ratio	P ₁ (psi)	P ₂ (psi)	P ₃ (psi)	P ₄ (psi)	P ₅ (psi)	P ₆ (psi)	P ₇ (psi)	P ₈ (psi)
1	3	3.915	50	195.7	766.3	3000	-	-	-	-
2	4	2.783	50	139.2	387.3	1077.9	3000	-	-	-
3	5	2.268	50	113.4	257.2	583.3	1322.8	3000	-	-
4	6	1.979	50	98.9	195.7	387.3	766.3	1516.2	3000	-
5	7	1.795	50	89.7	161.1	289.1	518.9	931.3	1671.5	3000

In designing the symmetrical wobble plate compressor, the optimum number of stages will choose by combination all number of stages that had been analysis

Table 2. Pressure ratio and pressure each stages.

Piston	Max force for 3 Stage (N)	Max force for 4 Stage (N)	Max force for 5 Stage (N)	Max force for 6 Stage (N)	Max force for 7 Stage (N)
1	3,454.1	3,692.434	933.982	1,928.590	1,470.029
2	4,110.9	4,590.086	1,111.590	2,229.694	1,664.679
3	4,892.7	5,705.961	1,322.974	2,577.808	1,885.103
4	-	7,093.113	1,574.554	2,980.272	2,134.714
5	-	-	1,873.976	3,445.572	2,417.376
6	-	-	-	3,983.517	2,737.466
7	-	-	-	-	3,099.941

Table 2 shows that the maximum force at each piston. and the force is directly proportional to the pressure in compressor.

This section will discuss about the optimization of the symmetrical wobble plate compressor design. As stated before the number of stages are between 3 until 7 for better result in design. It could be prove that by the result of pressure ratio. It becomes higher if we choose 1 or 2 stages. higher pressure ratio will decrease the compressor performance. If the number of stages more than 7, the compressor efficiency will increase but a dimension of compressor become bigger, production cost more higher and the compressor is no longer compact.

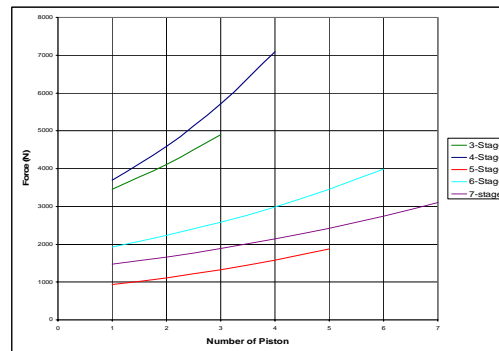


Fig.6 Load in each piston for each number of stages

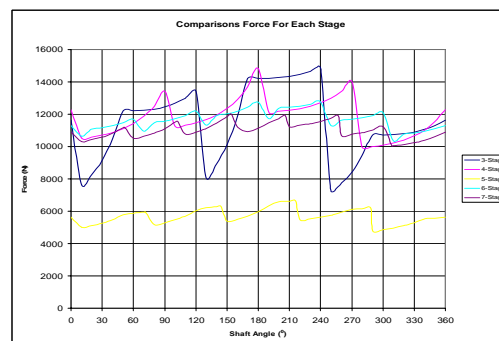


Fig.7 Compressor total force in each shaft angle rotation with the number stage of compressor

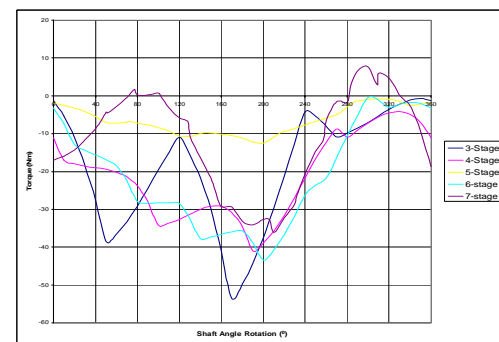


Fig.8 Total torque pada compressor setiap shaft angle rotation dengan number stage of compressor

Based on the force that has by each piston, piston number 5 has the optimum force. In Fig 6 and 7 it can be seen or determined the number of stages that more optimum. Torque in 3 stage of compressor has the bigger torque and minimum torque in stage 5. Usually, if the pressure ratio become smaller the torque is also small. But in case it is not happen because diameter of wobbles plate is big. The decrease of pressure ratio and diameter piston not proportional with the enlargement of the diameter wobbles plate as shown in Fig 9. In Fig 9 it can see that if the number of stages become bigger, the dimension of wobble plate also bigger and not for the diameter of piston, where it becomes smaller.

The diameter of piston in the 6th and 7th stages was bigger than in the 1st stage. If in the 6th and 7th stages the diameter of the 1st piston smaller than the 5th piston, the last piston was the smallest. It is impossible for process

production and the availability of groove to give the piston ring and raider ring or guide ring. If the number of stages become bigger so that the space that was needed for the cylinder also bigger. It is impossible to reduce the wobble plate radius. In this analysis the dimension for the size of parts already maximum for each stage.

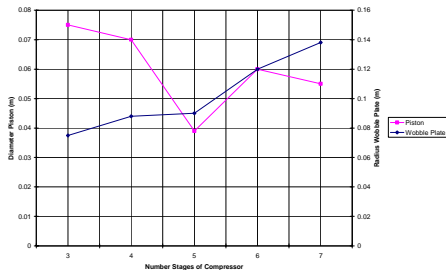


Fig.9 Hubungan diameter piston, radius wobble plate, dan number of stage of compressor

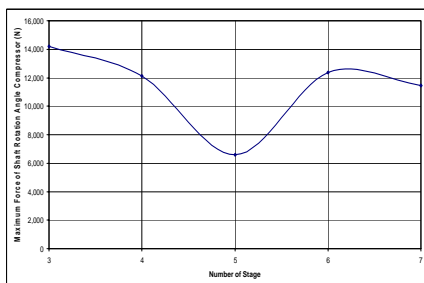


Fig.10 Force maximum on the compressor

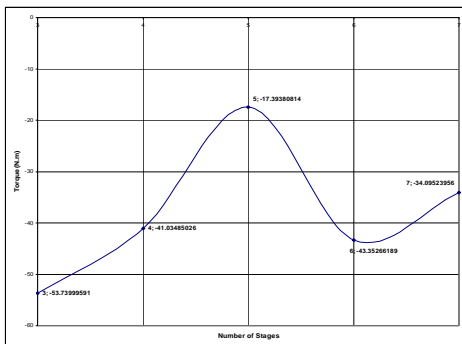


Fig.11 Torque maximum on the compressor

Fig 11 shows the torque in each of stage of compressor. It can see that this 5 stage produce smallest torque in the compressor. Before choosing the number of stages it is better to check the torque in compressor. Smallest torque will produce better performance compressor. This is to the energy or the work that done by compressor. In Fig 12, the toque becomes smaller, and works of the compressor also become smaller. So, the efficiency of compressor increasing and therefore this compressor will have better performance. The compressor torque also has relationship with the prime mover (motor).

Motor with smaller torque will have smaller geometry and therefore cheaper to buy. Therefore, this kind of compressor is suitable for home refueling and mini station. Another factor is that the total of geometry of compressor

system must also the driver include with the drive. The whole geometry must be small and compact.

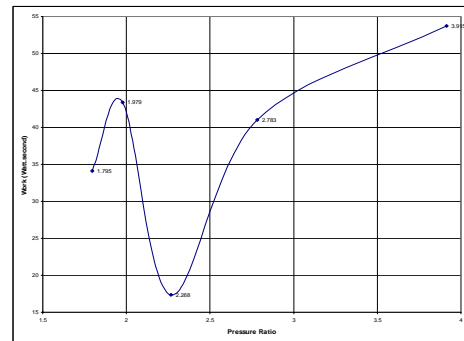


Fig.12 Work of compressor vs pressure ratio

6 Conclusion

Based on the analysis, we conclude that the optimum number of stages for capacity 10 m³/hr is 5. the specifications of the compressor are as follow:

Table 3. Optimum specification of symmetrical wobble plate compressor

Input Data	Reference Value				
	1 Stage	2 Stage	3 Stage	4 Stage	5 Stage
Cylinder diameter	39 mm	28.25 mm	20.47 mm	14.83 mm	10.74 mm
Suction Pressure	50 psi	113.4 psi	257.176 psi	583.26 psi	1322.79 psi
Discharge Pressure	113.4 psi	257.176 psi	583.26 psi	1322.79 psi	3000 psi
Stroke	47.96 mm				
Fluid	Natural gas				
Rotating Speed	1500 rpm				
Tilting Angle	16°				
Pressure ratio	2.698				
Capacity	10 Nm ³ /hr				

Reference

- [1] Ardiyansyah Syahrom., Md. Nor Musa., and Ainullofti Adb Latif. Wobble Design and Development of a Compressor for Natural Gas Vehicle Refuelling. *International Conference Engineering Design 2005*. August 15-18, 2005. Melbourne: ICED. 2005.
- [2] Ardiyansyah Syahrom., Md. Nor Musa., and Ainullofti Adb Latif. Wobble Plate Compressor for Natural Gas Vehicle Refueling. *Asia Pacific Natural Gas Vehicles Associations Conference & Exhibition*. July 26-28, 2005. Kuala Lumpur: ANGVA. 2005. 40.
- [3] Ardiyansyah Syahrom., Md. Nor Musa., and Wan Ali Wan Mat. Pembangunan Pemampat Simetri Berperingkat Jenis Plat Wobal Untuk Gas Asli. *Malaysian Science and Technology Congress*. September 23-25, 2003. Kuala Lumpur: COSTAM. 2003. 237.
- [4] A. Iongo Giovanni., and Gasparella Andrea (2003). Unsteady state analysis of the compression cycle of a hermetic reciprocating compressor. *International journal of refrigeration* 26.

- [5] Eastop. T.D., and McConkey. (1995). *Applied Thermodynamics For Engineering Technologists*. 5th ed. New York: John Wiley & sons, INC.
- [6] Harvey Nix. (2001). Compressor Analysis. In: Paul C. Hanlon. *Compressor Handbook*. New York: McGraw-Hill. 5.1-5.34.
- [7] Hiroshi Toyada., and Masaharu Hiraga. (1990). Historical Review of The Wobble Plate and Scroll Type Compressors. *SAE Congress Paper*.
- [8] Kenji Tojo, Kunihiko Takao, Youzou Nakamura, kenichi Kawasima and Yukio Takahashi. (1988). A Study on The Kinematics of A Variable Displacement Compressor For Automotive Air Conditioning. *1988 International Compressor Engineering Conference at Purdue*. July 18-21-1988. West Lafayette: Purdue University. 496-504.
- [9] Roycas N. Brown (1986). *Compressor Selection and Sizing*. Houston: Gulf Publishing Company.
- [10] Shimizu Shigemi., Shimizu Hidehiko., and Terauchi Kiyoshi (1989). *Wobble plate type compressor*. (US4869651).
- [11] Takahiro Nishikawa., hiroshi Nishikawa., Tomio Obokata., and Tsuneaki Ishima. (2000). A Study for Improvement on High Pressure Multistage Reciprocating Compressor. *International Compressor Engineering Conference at Purdue University*. July 25-28, 2000. West Lafayette: Purdue University. 105-112.
- [12] Woolatt Derek. (2001). Compressor Theory. In: Paul C. Hanlon. *Compressor Handbook*. New York: McGraw-Hill. 1.1-1.15.
- [13] Y.-C. Ma., and O.-K. Min (2001). Pressure Calculation in Compressor Cylinder by A Modified New Helmholtz Modeling. *Journal of sound and vibration*. 243(5): 775-776.
- [14] Zair Asrar Ahmad., Ardiyansyah Syahrom., Md. Nor Musa., and Ainullofti Adb Latif (2003). Kajian Tegasan Pada Pemampat Plat-Wobal Simetri Baru. *Malaysian Science and Technology Congress*. September 23-25, 2003. Kuala Lumpur: COSTAM. 2003. 237

Ratio Control of an Electromechanical Dual Acting Pulley – Continuously Variable Transmission (EMDAP- CVT) System Using PD-Fuzzy Logic Controller

Bambang Supriyo¹, Kamarul Baharin Tawi, Hishamuddin Jamaluddin and Sugeng Ariyono²

Faculty of Mechanical Engineering, Universiti Teknologi Malaysia
81310 UTM Skudai, Johore, Malaysia

¹Department of Electronic Engineering, Politeknik Negeri Semarang
Jl. Prof.H. Sudarto, S.H. Tembalang, Semarang 50329, Indonesia
E-MAIL : bsupriyo7763@yahoo.com

²Department of Mechanical Engineering, Politeknik Negeri Semarang
Jl. Prof.H. Sudarto, S.H. Tembalang, Semarang 50329, Indonesia

Abstract

Currently, a hydraulic system has been widely used in automotive application for changing transmission ratio of metal pushing V-belt Continuously Variable Transmission (CVT) utilizing a single acting pulley system. This system needs continuous power to supply force to both movable pulleys to maintain the desired transmission ratio and clamping force. The application of continuous hydraulic force introduces power loss. This paper introduces an alternative way of changing transmission ratio using an electromechanical system applied to dual acting pulley CVT system. This system has a viable solution for reducing the power consumption, since it operates only during changing the transmission ratio. The dual acting pulley system incorporates two movable pulley sheaves in both primary and secondary shafts to keep the belt position always aligned in the center of shafts, even during the ratio change, hence eliminating the belt misalignment effects. The electromechanical system introduces two DC motors as actuators and power screw mechanisms as mechanical linkages to adjust and to brake the axial position of movable pulleys. The primary motor is used for changing the transmission ratio, while the secondary one is used for preventing the belt from slipping. This paper proposes proportional differential (PD) fuzzy logic controller to control both primary and secondary DC motors such that both primary and secondary pulley radii satisfy the desired transmission ratio. The initial value of the PD gain is derived using Astrom and Hagglund tuning method. Then, this gain is fine tuned using fuzzy logic to improve the performance of the transmission ratio control automatically until it converges to a predetermined optimal global criterion. Computer simulation results are presented to demonstrate the effectiveness of the proposed controller.

Keywords :

Electromechanical CVT, Dual Acting Pulley CVT, Ratio Control, Fuzzy Logic Controller.

1. Introduction

Due to its wide range coverage, continuously variable transmission (CVT) enables the engine to run in its fuel-efficient operating point for any vehicle load, hence lowering the engine fuel consumption. Most current metal pushing V-belt CVTs are hydraulically actuated. This hydraulic system requires continuous power to supply clamping force. This continuous energy consumption becomes the major loss hence reducing CVT efficiency [1]. Although some efforts on improving the performance of hydraulic systems have been done using a new servo hydraulic actuation system [2],

a new electromechanically actuated CVT has been introduced in [3] to improve CVT efficiency.

Slightly different with [3], this paper introduces an electromechanical CVT with dual acting pulley system. This system adopts two movable pulley sheaves in each shaft to eliminate belt misalignment effects. Long term application of this misalignment may damage the belt. The metal belt misalignment has been studied intensively in [4]. Two DC servomotors are used as actuators. These motors will be activated when transmission ratio is changed, hence cutting the actuator operation time and reducing the energy loss.

PID (Proportional, Integral and Derivative) controller has been the basis in the simple linear control systems. The PID controller is a well-known technique for various industrial control applications, mainly due to its simple design, straightforward parameters' tuning, and robust performance. PID controllers are usually used to control servomotors. Position controls utilizing PID can be seen in [5],[6],[7],[8]. The conventional approach to determine the PID parameters is to use a mathematical model of the process and a simple tuning law with a fixed set of gain parameters. One famous example of such approach is the Ziegler-Nichols method [9]. However, an accurate model of a real industrial process is difficult to obtain due to its complex characteristics and parameter variations. Therefore, tuning laws based on these inaccurate models are no longer adequate.

In order to increase the capabilities of the PID controller in improving its performance, a fuzzy logic controller (FLC) will be used. The FLC presents an algorithm for translating an expert knowledge-based linguistic control strategy into an automatic control strategy [10]. Fuzzy-logic-based PID controllers offer knowledge representation and learning capabilities to tackle a wide range of complex dynamical systems, which may be ill defined or subjected to numerous varying parameters. An auto tuning of PD controller using Takagi-Sugeno type fuzzy logic controller is proposed for performing the ratio control of the EMDAP CVT system.

2. Background of CVT

The idea behind a continuously variable transmission is based on the use of a variator instead of gears. The variator consists of a primary and a secondary pulleys and a metal belt. The variator geometry is given in figure 1. By assuming that no gross belt slip occurs, the belt is inextensible and it runs at perfect circles with radii R_p (on the primary pulley) and R_s (on the secondary pulley), the tangential velocities of primary pulley, secondary pulley, and the belt will be the same.

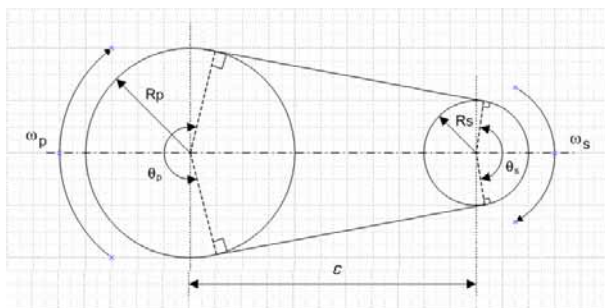


Figure.1. Variator Geometry

Based on this variator geometry, the relationship between speed and pulley radii can be given as follows:

$$\alpha_s R_s = \alpha_p R_p \quad (1)$$

$$r_s = \frac{\alpha_s}{\omega_p} \quad (2)$$

$$r_g = \frac{R_p}{R_s} \quad (3)$$

Where R_p , R_s , ω_p , ω_s , r_s , r_g are belt contact radius on the secondary pulley, belt contact radius on the primary pulley, angular speed of primary pulley, angular speed of secondary pulley, speed ratio, and geometrical ratio respectively. The following implicit relationship between belt length (L) and the pulley radii (R_p , R_s) can be derived as:

$$L = (\pi + 2\theta)R_p + (\pi - 2\theta)R_s + 2c \cdot \cos \theta \quad (4)$$

$$\theta = \arccos \left(\frac{R_p - R_s}{c} \right) \quad (5)$$

where c is the pulley center distance, and θ is the belt wrapped angle on the primary pulley.

2.1. Metal V-Belt CVT

A metal V-belt developed by Van Doorne's Transmissie of Holland consists of thin, flat tension bands (sometime called as rings) and metal V-segments (sometime called as blocks), which filled the entire length of the band. Each band is approximately 0.2 mm thick, and is laminated. Thin segments are strung together by two sets of thin bands positioned in slots at each side of the segments. One set of the band consists of about nine bands, and there are about 300 segments for each belt [11]. The metal belt structure [12] can be seen in figure 2.

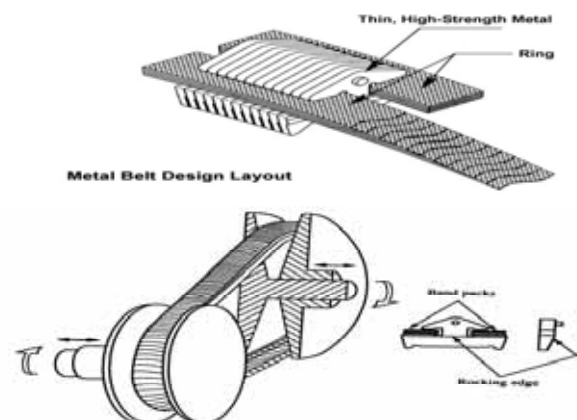


Figure 2. The structure of a Metal V-belt CVT (Srivastava and Haque, 2004)

2.2. Electromechanical Dual Acting Pulley CVT

The electromechanical dual acting pulley CVT (EMDAP CVT) system utilizes two DC servomotors as actuators. The system consists of two sets of pulleys, namely primary pulley placed on input fixed shaft, and secondary pulley placed on secondary fixed shaft. Each set of pulley has two movable sheaves that can be shifted axially along the shaft. The servomotor shaft is directly connected to a series of gear reducers and power screw mechanism to axially move the pulley sheaves. A Van Doorne's metal pushing V-belt is placed between pulley sheaves, and runs on the surface of the sheaves. This metal belt connects the primary and secondary pulleys to transmit the power and torque from the input to the output shaft by means of friction between the belt and the pulley contact [13],[14].

The contact radius of the belt with primary pulley (R_p) and the contact radius of the belt with secondary pulley (R_s) determine the transmission ratio. The primary and secondary parts of the CVT system have the same components. The components of each part (primary or secondary) consist of a dc servo motor, a gear reducer with ratio of (30:1), a gear reducer with gear ratio of (60:14), power screw mechanism, and two movable metal pulley sheaves for clamping the metal belt. A spring disc is inserted in the back of each secondary pulley sheave to provide continuous clamping force to the belt, and to reduce excessive slip during transmission ratio change. The block diagram of the EMDAP-CVT system can be seen in figure 3.

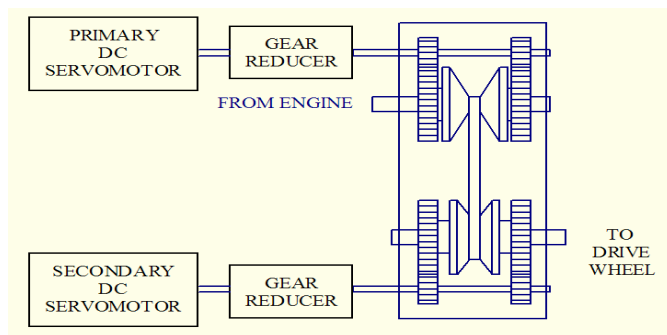


Figure 3. The block diagram of the EMDAP-CVT

3. System Modeling

3.1. Modeling of DC Servomotor

The dynamic model of the DC servomotor is represented as follow:

$$V_a = sL_a i_a + R_a i_a + K_a \dot{\theta} \quad (6)$$

$$V_a - K_a \dot{\theta} = s.L_a .i_a + R_a .i_a \quad (7)$$

$$T_m - T_L = J_m .s.\omega + B_m .s^2 .\omega \quad (8)$$

$$\theta = \omega/s \quad (9)$$

Where :

V_a : Motor Voltage [V]

L_a : Motor Inductance [H]

i_a : Motor Current [A]

R_a : Motor Resistance [Ω]

K_a : Back emf constant [mV/(rad/sec)]

ω : Motor shaft angular velocity [rad/sec]

θ : angular displacement [rad]

T_m : Motor Torque [Nm]

K_m : Torque Contant [Nm/A]

T_L : Load Torque [Nm]

J_m : Motor Inertia [Nm.sec²]

B_m : Viscous friction coefficient [Nm/rad/sec]

3.2. Gear Reducer and Power screw Mechanism

A gear reducer serves as a speed reducer and torque multiplier. The gear reducers are coupled with the servomotor shaft. The gear reducer output is connected to power screw mechanism for shifting the pulley sheaves. There are two gear reducers employed in the EMDAP-CVT system. The first one has ratio of (30:1), while the second one has the ratio of (60:14). The total gear ratio of these reducers is (128.57:1). The power screw mechanism is used to move the pulley sheaves axially. It converts every 360° of rotation into 2 millimeters axial movement. The combination of gear reducers and power screw mechanism is used to help the motor in providing significant torque to turn the power screw mechanism.

4. Proposed Controller

This paper proposes a Fuzzy-PD controller for each DC servomotor system to control the transmission ratio of the CVT. The primary DC servomotor system will tracks the desired value of the primary pulley radius, and the secondary DC servomotor system will take the actual value of the primary pulley radius as its input and set the secondary pulley radius accordingly to prevent belt slip. The common problem of tuning PID is dealing with specifying three gain parameters to meet a certain requirement of control specifications, such as set-point following, load disturbance attenuation, robustness due to model uncertainties, and noise rejection [15]. Many different techniques have been proposed in the literature to cope with these tuning problems. The use of Ziegler-Nichols formula for tuning the PID parameters usually gives a good result in load disturbance attenuation, but poor in both lowering the overshoot and reducing the settling time. These two aspects are mainly triggered by the selection of high proportional gain. To cope with this problem, a fuzzy logic controller is proposed to obtain the suitable weight value to adjust the proportional gain with respect to the current values of the system error (e) and its change of error rate (de). The block

diagram of the proposed controller scheme is given in the figure 4.

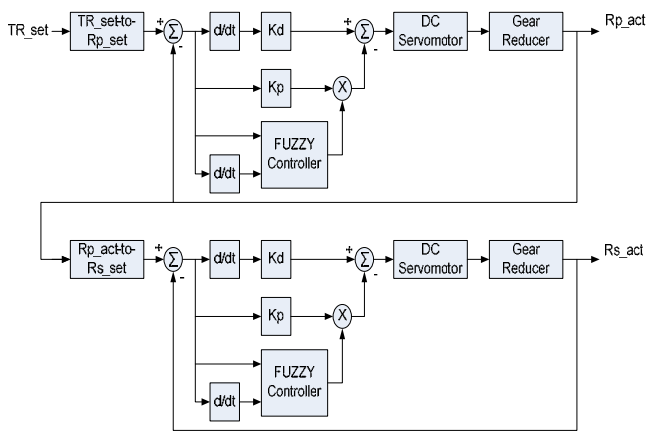


Figure 4. Block diagram of the Fuzzy-PD controller

4.1. Parameter Tuning for PID Controller

The transfer function of a PID controller has the following form:

$$G_{PID} = K_p + K_i / s + K_d s \tag{10}$$

$$G_{PID} = K_p (1 + 1/(T_i s) + T_d s) \tag{11}$$

where K_p , K_i , and K_d are the proportional, integral, and derivative gains, respectively, $T_i = K_p / K_i$ and $T_d = K_d / K_p$. T_i and T_d are the integral time constant and the derivative time constant, respectively.

In order to obtain initial parameters of PID controller, the Astrom-Hagglund method [16] will be used to determine both critical period of waveform oscillation (T_c) and critical gain (K_c). These two values can be obtained by running the closed loop control of DC servomotor system with relay feedback controller. The oscillation period of the output waveform is considered as the critical period attained from a proportional feedback. The critical gain is given as follows:

$$K_c = \frac{4d}{\pi a} \tag{12}$$

Where d is the amplitude of the relay output, and a is the amplitude of the waveform oscillation. Based on the values of T_c and K_c , the PID parameters (K_p , T_i , and T_d) can be specified using Ziegler-Nichols formula [9] (see Table 1.).

Table 1. Ziegler-Nichols parameter tuning.

	K_p	T_i	T_d
P	$0.5 K_c$		
PI	$0.45 K_c$	$0.85 T_c$	
PID	$0.6 K_c$	$0.5 T_c$	$0.125 T_c$

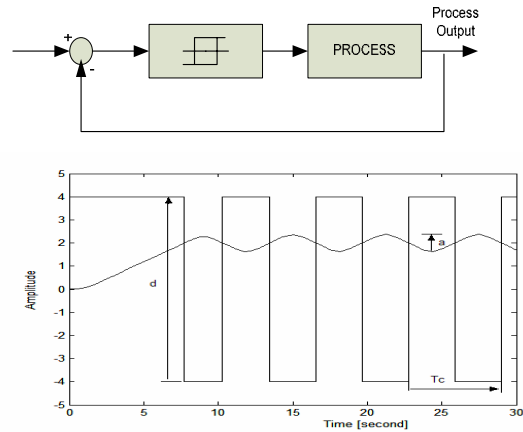
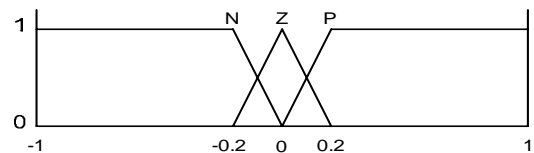
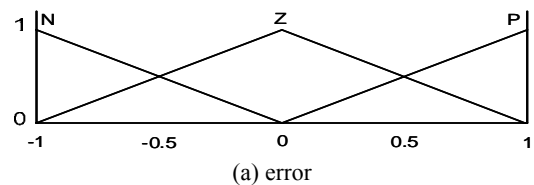


Figure 5 The relay feedback controller

4.2. Fuzzy Auto Tuning

Fuzzy logic is used to perform online tuning for PID parameters. A tuning algorithm proposed in this paper only tunes the proportional gain (K_p) of the PID controller. The fuzzy auto-tuner has two inputs: the error (e) and the change error rate (de) of the output response. To normalize these inputs into the range of $[-1,1]$, these inputs are scaled by two coefficients, K_e and K_{de} respectively. The fuzzy system can be started to operate at a certain error value by specifying the value of K_e . This paper selects $K_e = 10$, meaning that the system will effectively operate when the output error is in the range of $[-0.1, 0.1]$. The K_{de} can be set initially to '1'. Three triangular membership functions (see figure 6) are defined for each input (N = Negative, Z = Zero, P = Positive), while five constant values of Takagi-Sugeno type in the range of $[-1,1]$ is defined for the output (NB = Negative Big, NS = Negative Small, Z = Zero, PS = Positive Small, PB = Positive Big).



(b) the change of the error

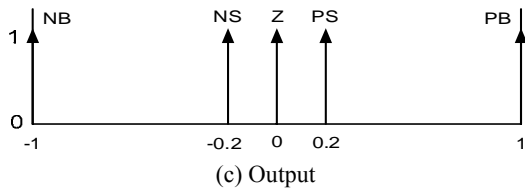


Figure 6. Membership function of input and output

The error ($e(k)$) and the error change rate ($de(k)$) are defined as follow:

$$e(k) = x_r(k) - x_a(k) \tag{13}$$

$$de(k) = e(k) - e(k-1) \tag{14}$$

The fuzzy output (Y_f) is the singleton value in the range of $[-1, 1]$ of the Takagi-Sugeno Type. This output value is then multiplied by the proportional gain (K_p) and then summed with the derivative gain to give the overall controller gain (G_c):

$$G_c = Y_f * K_p + K_d \tag{15}$$

The fuzzy rules to determine the output fuzzy (Y_f) is the following equation:

$$R^i: \text{if } e(k) \text{ is } A_j \text{ and } de(k) \text{ is } B_j \text{ then } Y_{fi} \tag{16}$$

Where $i (i=1, \dots, 9)$ is the number of rules.

The rule justification is based on the “scale mapping” method developed by King and Mamdani [10]. This justification is done by referring to a closed system trajectory in a phase plane. The control rules of the consequence of (Y_f) are determined from the error curve of the system step response [9] (see figure 7). This justification requires both knowledge of controller parameter adjustment based on the phase plane analysis, such as overshoot and rise time, and intuitive feeling of the behavior of the closed loop system. The prototype of the fuzzy control rules is given in table 2. While the justification of a fuzzy control rules is presented in table 3.

Defuzzification is needed for the fuzzy logic controller to convert its internal fuzzy output variables into crisp values usable for the controlled system. This paper will adopt the weighted average method represented by:

$$y_f = \frac{\sum_{i=1}^n w_i c_{j,i}}{\sum_{i=1}^n w_i} \tag{17}$$

Where :

n is the number rules, $C_{j,i}$ is the vale of the center of gravity of the fitness of i -th rule in the premise, and the fitness (w_i) is defined as:

$$w_i = \mu_{A_j}(e(k)) \times \mu_{B_j}(de(k)) \tag{18}$$

$$j=1,2,3, \quad i=1,2,3$$

This type of defuzzification method is selected due to its simple and fast computation [8].

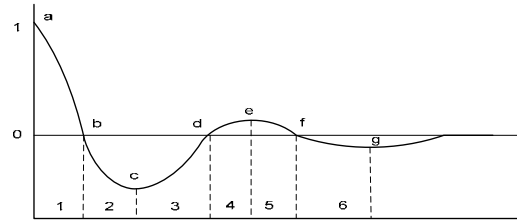


Figure 7. Error curve of output

Table 2. The prototype of the fuzzy control rules

Rule No.	e	de	Y_f	Reference Point
1	P	Z	PB	a,e
2	Z	N	NS	b
3	N	Z	NB	c
4	Z	P	PS	d
5	Z	Z	Z	f

Table 3. the justification of a fuzzy control rules

Rule No.	e	de	Y_f	Reference range
6	P	N	PB	1,5
7	N	N	NB	2
8	N	P	NB	3,6
9	P	P	PB	4

5. Simulation Studies

Simulation studies of the proposed Fuzzy-PD controller are carried out in order to investigate its effectiveness in this ratio control application. In these studies, the DC servomotor has the following important parameters as shown in Table 4.

Table 4. DC servomotor parameters

Parameters	Values
Motor Voltage	24 V
Armature Resistance	0.03 Ω
Armature Inductance	0.1 mH
Back emf Constant	7 mV/rpm
Torque Constant	0.0674 N/A
Nominal Torque	1.6 Nm
Nominal Speed	3000 rpm
Rotor Inertia	0.1555e-4 N.sec ²

The simulation of the controller was performed using MATLAB/FUZZY-SIMULINK packages. Control performance is determined based on trajectory errors. An input excitation of sinusoidal waveform, is used to examine the performance of the conventional PD and the Fuzzy-PD controllers. The initial proportional gain will be set to the same value to ensure reasonable comparisons between these two controllers.

In order to obtain initial parameters of PID, the Astrom-Hagglund method based on a relay feedback controller is carried out to attain the critical period of waveform oscillation (T_c) and critical gain (K_c). The amplitude of the relay controller is set to 10 since the input voltage in the range of [-5,+5 volts] is needed to drive the servo system. From simulation results, these following parameters are found: $T_c = 0.4$ s, $a = 0.1$, and $d = 10$. By using equation (12), the critical gain (K_c) is 127.4. Finally, by using these values, Ziegler-Nichols formula, and equations (10) and (11), the three parameters of PID can be specified as follows: $K_p = 76.4$, $K_i = 636.9$, and $K_d = 3.8$. From these data, the integral gain (K_i) is too big and it is not suitable for this application, so the PID controller will be reduced to PD controller.

Figure 8 and figure 9 show the output responses of the sinusoidal wave excitation. From figure 9, it can be seen that the Fuzzy-PD controller has smaller position error (about -0.003 to +0.003) compared to the conventional PD controller (about -0.02 to +0.02). Thus, the Fuzzy controller has significantly improved the performance of the conventional PD controller, especially in terms of minimizing the position error of trajectory tracking.

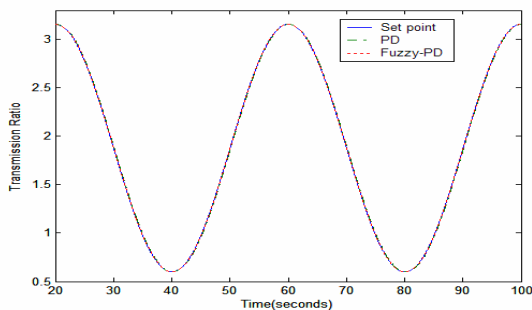


Figure 8. Set-point follower

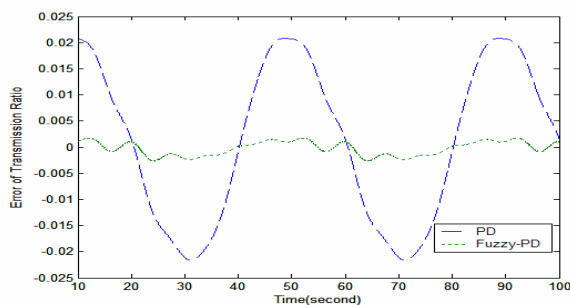


Figure 9. Position error of set-point follower

6. Conclusion

The conventional PID is initially tuned using Astrom-Hagglund method and Ziegler-Nichols formula. However, only both proportional and derivative gains of the PID are used to control the servomotor system. The integral gain is not used, since it is too big and unsuitable for this application. The proposed controller utilizes a fuzzy logic controller to tune the proportional gain of the PD controller and to improve the PD performance. The simulation results have shown that the Fuzzy controller has significantly improved the performance of the conventional PD controller, especially in terms of minimizing the position error of trajectory tracking for ratio control of the EMDAP-CVT.

Reference

- [1] Akerhurst, S. and N.D. Vaughan.1999. An Investigation into the Loss Mechanisms Associated with an Automotive Metal V-belt CVT . In *European Automotive Congress Vehicle Systems Technology for The Next Century*, STA991407, Barcelona.
- [2] Veenhuizen, P.A. and S. Cools. 2004. Servo Hydraulic Control of a Continuously Variable Transmission. In *7th International Symposium on Advanced Vehicle Control (AVEC'04)*, Arhem, Netherland.
- [3] Van de Meerakker, K., P. Rosielle, B.Bonsen and T. Klaassen. 2004. Design of An Electromechanical Ratio and Clamping Force Actuator for a Metal V-Belt Type CVT. In *7th International Symposium on Advanced Vehicle Control (AVEC'04)*, Arhem, Netherland.
- [4] Tawi, K.B. Investigation of Belt Misalignment Effects on Metal Pushing V-Belt Continuously Variable Transmission, Ph.D. Thesis, Cranfield University. 1997.
- [5] Shi Jingzhuo, Xu Dianguo, and Wang Zongpei, 2000. Hybrid Stepping Motor Position Servo System Based On DSP. In *IECON 2000, 20th Annual Conference of the IEEE Industrial Electronics Society IEEE*, Vol.2, 22-28 Oct 2000, pp.1475-1479.
- [6] Habib, R. Maki, 2001. Designing Fuzzy Logic Controllers fro DC Servomotors Supported by Fuzzy Logic Control Development Environment. In *IECON'01: The 27th Annual; Conference of the IEEE Industrial Electronics Society*.
- [7] Yu Yang, Wen-Ge Wang, De-Jie Yu, nd Ge Ding, 2002. A Fuzzy Parameters Adaptive PID Controller Design of Digital Positional Servo System. In *Proc. Of The First International Conference on Machine Learning and Cybernetics*. Beijing.
- [8] Yusuf I. Bulale, M.J.E. Salami, Marizan Sulaiman, 2003. Design and Development of DSP-based hybrid Controller For Servo Driver Applications. In

- Proceeding of 2003 IEEE Conference on Control Applications*. Vol.1, 23-25,pp.773-778.
- [9] Hyung-Soo Hwang, Jeng Nae Choi, Won Hyok Lee, Jin-Kwon Kim, 1999. In *A Tuning for The PID Controller Utilizing Fuzzy Theory*, "IJCNN'99 International Joint Conference on Neural Network". Vol.4,10-16,pp. 2210-2215.
- [10] Lee, C.C, 1990. Fuzzy Logic in Control Systems: Fuzzy Logic Controller- Part I. *IEEE on Trans. On System, Man, and Cybernetics*. Vol.20. no.2.
- [11] Fujii, T. , T. Kurokawa and S. Kanehara. 1993. A Study of a Metal Pushing V-Belt CVT Type Part 1: Relation Between Transmitted Torque and Pulley Thrust, *SAE Paper No. 930666*.
- [12] Srivastava, N. and I. Haque. 2004. On the Operating Regime of a metal pushing V-belt CVT under steady state microslip conditions. In *2004 International Continuously Variable Transmission and Hybrid Transmission Congress*, September 23-25.
- [13] Kanehara, S. , T. Fujii and S. Oono.1996. A Study of a Metal Pushing V-Belt Type CVT : Macroscopic Consideration for Coefficient of Friction between Belt and Pulley, *SAE Paper No. 9636277*.
- [14] Fushimi, Y. , T. Fujii and S. Kanehara. 1996. A Numerical Approach to Analyse the Power Transmitting Mechanisms of a Metal Pushing V-Belt Type CVT. *SAE Technical Paper Series 960720*.
- [15] Visioli, A.,1999. Fuzzy Logic Based Set-Point Weight Tuning of PID Controllers. *IEEE Trans. On Systems, Mans, and Cybernetics-Part A: System and Humans*, Vol.29, No.6.
- [16] K.J. Astrom and T. Hagglund, 1984. Automatic Tuning of Simple regulators with Specifications on phase and amplitude margins. *Automatica*, 20,pp.645-651.

Micro/Meso Mechanical Manufacturing (M^4): Opportunities and Challenges - A Review

R. Mehfuz, M. Y. Ali*

Department of Manufacturing and Materials Engineering
Faculty of Engineering, International Islamic University Malaysia
53100, Jalan Gombak, Selangor Darul Ehsan, Malaysia

Abstract

Besides conventional precision manufacturing and micro-electro-mechanical system (MEMS), micro/meso mechanical manufacturing (M^4) is emerging as the most potential contender to fabricate complex micro/meso-scale components using a variety of engineering materials. M^4 domain covers the components sizes from $10\ \mu\text{m}$ - $10\ \text{mm}$ which is inconvenient to fabricate by other processes. The inclusion of 'microfactory' concept in producing micro/meso-scale parts shows huge prospects in terms of low power consumption, reduced space requirement and higher production. This paper presents the review of M^4 . It includes the recent effort, micro-end-milling, microfactory, challenges and opportunities.

Keywords: Micro/Meso mechanical machining (M^4), MEMS, precision machining, microfactory.

1. Introduction

Miniaturized components are increasingly in demand for numerous applications in diversified fields, like aerospace, automotive, biomedical, healthcare, electronics, environmental, communications, and consumer products. Researchers have been working on the microsystems that promise to enhance health care, quality of life and economic growth in such applications as micro-channels for lab-on-chips, shape memory alloy 'stents', fluidic graphite channels for fuel cell applications, subminiature actuators and sensors, and medical devices [1]. Thus, miniaturization technologies are perceived as potential key technologies of the future that will bring a completely different approach of man-machine interaction with the physical world.

Currently, common techniques utilized in the fabrication of micro-components are based on the techniques developed for the silicon processing, the base of micro-electro mechanical system (MEMS). The boom of MEMS technologies in recent years would not have been possible without the maturity of IC industry, as these two do share many common fabrication processes. Unfortunately these processes are cost effective only in large volume and are limited to production of simple planar geometries. It is also limited to two dimensional (2D) and pseudo three dimensional (2.5D) features in a narrow range of material, usually silicon-based [2.] These lithographic techniques are virtually unable to produce 3D curved features and are strictly sensitive to operating environments and high cost. Nontraditional fabrication methods, such as focused ion beam machining, laser machining, and electro-

discharge machining, are capable of producing high-precision micro-components, but have limited potential as mass production techniques due to the high initial cost, poor productivity, and limited material selection [3]. All these issues limit the range of MEMS application whose demand is increasing day by day. Research into the development of flexible, effective production technologies for micro components has therefore intensified.

Conversely, the conventional precision mechanical machining techniques is quite capable of producing 3D complex micro range features of varieties of materials with good accuracy. They are largely applied in watch making, optics apparatus and in some electronic manufacturing. However, there operation environment is extremely strict. The costs of production and initial investment are too high. The machine size is also large when compared to ordinary machine tools, despite the workpiece being much smaller than ordinary workpieces. Thus high power consumes to produce such small parts. Hence, this kind of equipment is not the most optimal tools to fabricate micro features. In order to overcome the aforementioned limitations of the existing technologies for miniaturization, a new complementary approach of manufacturing is needed to be emerged. To meet up this gap micro/meso mechanical manufacturing (M^4) aroused great interest among the researchers. This technique is suitable to fabricate $10\ \mu\text{m}$ - $10\ \text{mm}$ sized microcomponents.

2. M^4 - The Missing Link

The current technological capabilities are expressed in terms of relative accuracy versus the manufactured feature size as shown in Figure 1. It is seen that there exists a domain, approximately from $10\ \mu\text{m}$ - $10\ \text{mm}$ characterized

* Corresponding Author. E-mail: mmyali@iiu.edu.my
Tel: +603-6196-4507, Fax: + 603-61964477

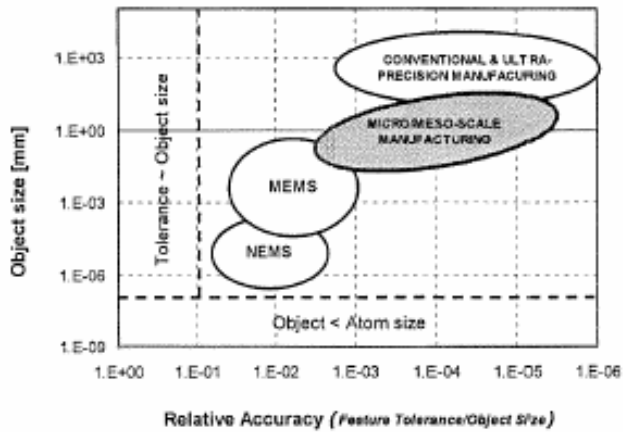


Figure 1. Micro/meso scale manufacturing domain [4]

by relative accuracies and precision of 10^{-3} to 10^{-5} . For this domain no specially developed manufacturing technologies are in existence [4]. This crucial domain requires the development of new classes of manufacturing processes which is defined as M^4 .

Thus M^4 is considered as the missing link between conventional precision manufacturing and MEMS. The motivation of M^4 basically comes from the translation of macro-process knowledge to the micro-level. There is evidence to suggest that M^4 is technically better for some cases than the currently used MEMS techniques [2]. Thus, miniaturized machine tools and tiny cutters are introduced. The concept of the microfactory has been proposed in 1990 and estimated the effect of miniaturization of production systems. This concept coincides with M^4 processes, which would not only provide considerable savings in terms of energy, space and resources but would also facilitate changes in the production line layout. It could change the long-established image of factories and would also enable a wide variety of production systems and various other future possibilities.

3. Opportunities of M^4

Compare to other manufacturing techniques, M^4 has its own distinctive advantages, such as high relative accuracy, individual feature control, complex 3D geometry fabrication, less material limitations, higher aspect ratio [5] etc. But the most important factors influencing the miniaturization through M^4 is the significant reduction of unit cost. Here cost can be reduced in varieties of fashions including capital cost of space, energy, operating, material costs etc. Because the size of the miniaturized machine tools is commonly a tenth to several tenths of the size of the conventional machine tools, the effect of power saving is very remarkable. It was estimated that in the case of a 1/10 size-reduction of production machines, the total energy consumption in the factory decreases to approximately 1/100 of that of a conventional factory [5]. The characteristic of miniaturized machine tools provides greater advantages for rapid manufacturing. Small-size products can be fabricated by on-site manufacturing system

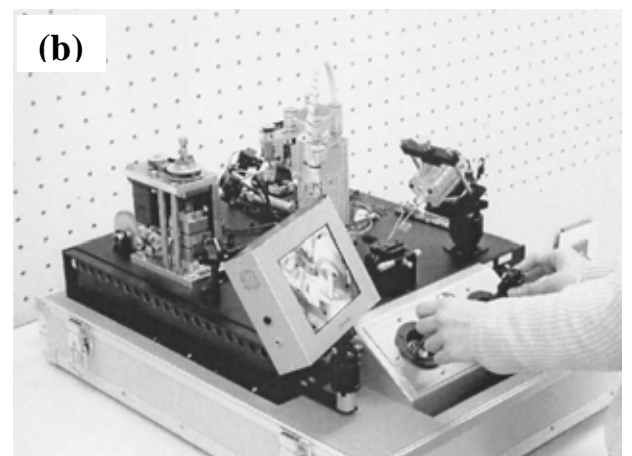
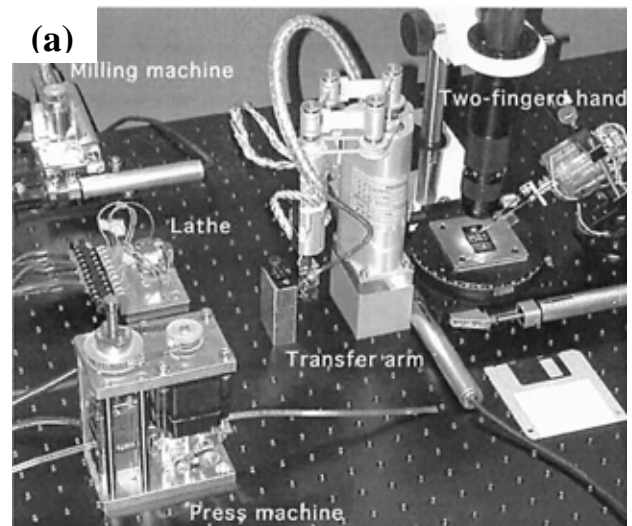


Figure 2. images of microfactories: a) Desktop microfactory, b) Portable microfactory [6].

at storefront, office or home. A mobile factory loading micro production system can make products while transporting between material supplier and customer. Thus requirements of a flexible manufacturing system can be satisfied by the convenient re-configurability. Desktop microfactory, as shown in Figure 2a, is composed of two areas within a 500 mm \times 700 mm space, i.e. the machining area and the assembly area. In the machining area, microparts are machined by a microlathe, micro-milling machine and micro-press machine. Machined parts are transported by a micro-transfer arm to the assembly area. Finally, a two-fingered microhand assembles the parts [6]. Figure 2b shows a portable microfactory that was designed to be packaged in a suitcase to demonstrate its portability. This portable microfactory has external dimensions of 625mm length, 490mm width and 380mm height and weighs 34 kg (main body weight: 23 kg). It can also be placed in the trunk of an automobile and transported with mounted casters [6].

Given the magnitude of the forces that are generated in micro/meso scale machining, there is no reason to believe that systems as massive as those that have been used in macro machining ultra precision components will be required to achieve stiffness and damping characteristics

that are necessary for meso-scale systems. Similar case is for the magnitude of relative motions between the tool and the part feature being created. Besides the reduced space and energy requirement micro-factory requires about one to two orders of magnitude lower in construction costs compared to conventional macro systems. Some Korean researchers, constructed a precision 5-axis milling machine of sizes 294mm x 220mm x 328mm (X, Y, Z directions) at a cost about 1/10th of the cost of macro sized precision milling machine available in the market [7].

In addition, micro-processes, micro-machines and micro-factories will reduce or eliminate the detrimental effects on environment by consuming very small amount of coolants/lubricants, dielectrics and electrolytes.

There are number of areas that can be benefited by the introduction of M^4 in the manufacturing realm. Such industries include electronics (e.g. small fasteners and connectors), optics (e.g. cameras, lenses) and medical (e.g. pumps, valves). Several bio micro-electro-mechanical systems (bio-MEMS) facilities are currently investigating ways to produce bio-MEMS devices based on the micro/meso mechanical processes of plastic micro-injection and hot embossing [8,9]. Other areas that can also be benefited by M^4 include defense (e.g. smart clothing and surveillance), space (light weight and smaller components, light weight production systems) and communications (e.g. optical switches) etc.

Several research and development activities in the micro/meso mechanical manufacturing have already been initiated in Japan, Taiwan, Europe (especially Germany) and USA focusing mainly on two aspects- one is designing and developing micro machine tools in order to verify its performance and the other is investigation of some distinctive phenomena in the M^4 technology and their mechanisms [7,10]. If we look upon the development of computer systems in the last four decades, it can be found that a remarkable reduction in size while at the same time performance increased by multiple order of magnitude [11]. It was only possible by the parallel development of new and supporting technologies and by shifting the paradigm into design and manufacture. The same development can be attained in the domain of M^4 . In the following section, the micro end milling is discussed in detail.

4. Micro End Milling

Recently micro-end milling process shows great interest among the researchers. Because, the geometries, that can be produced by micro-end-milling are more flexible than lithography based MEMS processing techniques. Moreover, a broad range of materials can also be processed by this. So far, several works had been conducted regarding various issues of micro-end-milling. In most of the studies on the M^4 technique were performed by using conventional-scale machine tools despite there have size effect on machining [12,13]. The relationship between micro-end-milling process parameters and cutting conditions has not been thoroughly discovered yet.

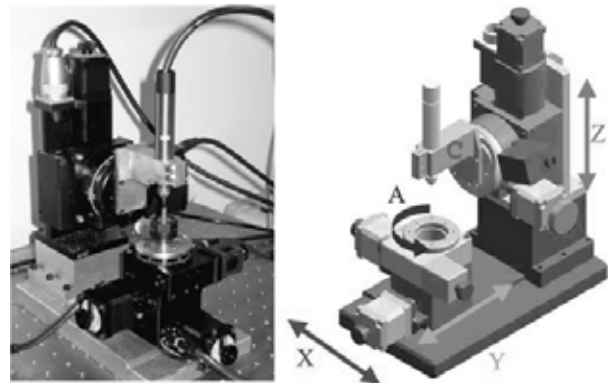


Figure 3. A 5-axis micro milling machine for machining micro parts [7].

4.1 Tool Life and Tool wear

Micro-end-milling on copper using 1mm diameter of tools showed that the tool life increases up to a certain extent with an increase in axial depth of cut. It was shown qualitatively that the relative cutting forces increased in conjunction with wear of the tool edges [13]. It was also found that tool wear in a micro-milling process is partly responsible for the stiffness of the micro-end-mill tool; because of the spring-back of an end mill under the thrust force, the effect of tool run out is reduced and the abrasion of each tool edge tends to be uniformity [14].

In micro-end-milling, premature tool failure is mostly happened. Micro tools less than 1mm diameter breaks abruptly during machining, which causes subsequent surface damage of the workpiece. To predict the possible breakage of the tool in advance and to extend the tool life, a smart workpiece holder (SWH) was introduced, which monitors the cutting force and reduces the metal removal rate when it predicts tool breakage [15]. Piezo-electric actuator was used to develop SWH and experiment is performed on mild steel and aluminium workpieces while in more than 50% cases tool breakage was predicted rightly and in 30% cases tool life was properly extended.

Turning operation [16], increase of cutting forces with the wear of tool edges was observed in micro-end milling experiments of copper [13, 17]. Cutting force model for micro-end-milling operation is developed by calculating the chip thickness while the effect of feed per mill tooth to tool radius (f_z/r) is considered [18], considering the effect of tool run-out [19] and tool wear [17]. Machining with edge-radius tooling shows that chips are not formed for chip thickness values less than a critical value [20]. But the chip thickness is obtained different for different materials [21].

4.2 Surface Roughness

Micro-structure of steels plays important role in achieving accuracy in micro cutting [22, 23]. It was observed that the ferrite grains would 'spring back' more than the pearlite grains during a micro cutting process. Cutting process performance was found improved by reducing grain size after an appropriate heat treatment. The roughness of the machined surface was found deteriorate with increasing cutting speed.

Another research showed the formation of built-up-edge (BUE) and its effects on surface roughness by means of cutting various steel materials. It was observed that the surface roughness of dual-phase and multi-phase steels is worse than single phase materials [24]. This probably happens due to the formation of BUE. In a micro-cutting process, to ensure better surface finish and higher accuracy, it is necessary to cut at a lower thrust speed. If machining tolerance is desired to be less than $1\mu\text{m}$ the feed per tooth and the cutting depth should be maintained at less than $1\mu\text{m}$. Because, when the ratio is less than a critical value, a significant ploughing phenomenon will appear [25].

In a recent research on micro-end milling process, three factor full factorial experiment was performed to determine the impact of cutting speed, chip load (or feed per tooth) and depth of cut on surface roughness [26]. Chip load was found most dominant factor that affects surface roughness. Surface roughness deteriorates with the increase of chip load. The interaction between chip load and cutting velocity also affects the surface roughness. At high chip loads, the effect of cutting speed is proved much more prominent [26, 27]. Investigation on the burr formation in micro-end-milling process is also carried on [28]. The method of minimization of burr formation is also studied [29].

4.3 Vibration

Vibration adversely affects the surface roughness of the machined part as well the tool life. In case of micro-components the effect of vibration should be much in control. Not sufficient studied performed so far regarding the vibration in M^4 . Experiment shows that higher feed rate increase vibration which ultimately results poor surface quality [30]. Another study shows that vibration increases sharply from $0.2\mu\text{m/t}$ to $0.4\mu\text{m/t}$ while the range of vibration decreases rapidly from $1.2\mu\text{m/t}$ to $1.6\mu\text{m/t}$ feed rate [31].

5. Challenges for M^4

The M^4 shows huge interest among researchers due to its ability to fabricate complex objects with wide range of materials. But the full scientific understanding is not developed yet. At micro level, machining performance is highly influenced by small vibration and excessive forces. Researchers still does not pay attention to the machining stability problem sufficiently, which should be reliable for the practical application. The tools used in micro/meso-scale processes are much smaller in dimension than conventional tools, resulting in lower stiffness, which combine with the higher forced obtained due to the size effect of the process. Size effect, effect of microstructure, type of chip and chip thickness- all of these individually or in combination can enforce the tool and workpiece systems into a vibration condition which are needed to be examined and investigated. Regenerative chatter effects also need to be considered, but accurate dynamic measurement between the tool and the work piece is extremely challenging. The impact of tool geometry on machining is needed to be analyzed. Burr formation and its minimization methods in

micro-end-milling processes are studied recently but future work on burr prevention should focus more on tool design.

In short the tool/workpiece interaction for different situations including different materials, microstructure, tribology is needed to be explored. Because in micro-cutting, chips may not form when the depth of cut is less than a critical chip thickness. The minimum chip thickness is about 30% of the tool edge radius [32]. This is due to elastic-plastic deformation of the machined part when machining at very small chip thickness, which may also cause cutting instabilities. Machining behavior in different work materials should also need to be studied extensively including metal matrix composites.

To compete economically, it is necessary for the M^4 processes to produce large quantities at a reasonable time and cost. To meet up this commercial challenge, integration of automation and microfactory can be effective by producing huge quantity parts in parallel or by producing 'very quickly' in series. Japan, Germany & USA is leading in this endeavor. Due to the downscaling of the component size, the handling, assembling, and testing of the micro/meso-scale components is also difficult, very time consuming and costly. User friendly devices in this regard are needed to be available, so that the product can be commercially viable.

6. Conclusion

Miniaturization promises huge potential in improving quality of life through efficient use of resources, higher economic growth and so on. It initially uses the micromachining techniques which finally creates a gap between traditional macro machining and MEMS. M^4 is able to bridge the gap between these two domains. Moreover, the ability to produce complex 3D geometry of wide range of materials shows that the development of M^4 is necessary for both technical and commercial point of views. Among the M^4 techniques, micro end milling shows a great interest among the researchers worldwide. It uses the tool of about 1 mm diametere. Smart work piece holder is used to monitor the tool breakage and life. Machines that are some orders of magnitude smaller than the conventional machines can be more useful to perform this micro/meso scale machining. Extensive research on this potential area should be encouraged through government and non-government initiatives.

References:

- [1] Madou, M. J. 1997. *Fundamentals of Microfabrication*, CRC Press, Boca Raton.
- [2] Kussul, E. M, Rachkovskij, D. A., Baidyk T. N., and Talayev S.A. 1996. Micromechanical engineering: A basis for the low-cost manufacturing for mechanical microdevices using microequipment, *Journal of Micromechanical Microengineering*, 6: 410–425.

- [3] Masuzawa, T., and Tonshoff, H. K. 1997. Three-dimensional micromachining by machine tools. *Annals CIRP*, 46: 621–628.
- [4] Ehmann, K. F., DeVor R. E., and Kapoor S. G. 2002. Micro/meso-scale mechanical manufacturing-opportunities and challenges. In *Proceedings of JSME/ASME International Conference on Materials and Processing*. 1: 6-13.
- [5] Okzaki, Y., Mishima, N., and Ashida, K. 2002. Microfactory and micro machine tools', In *Proceedings of The 1st Korea-Japan Conference on Positioning Technology*, Korea.
- [6] Tanaka, M. 2001. Development of desktop machining microfactory' *RIKEN Review* No. 34.
- [7] Bang, Y. B, Lee, K. M, Oh, S. 2005. 5-axis micro milling machine for machining micro parts. *International Journal of Advanced Manufacturing Technology*. 25: 888-894
- [8] Howe, R. T. 2003 Micro Systems Research in Japan, World Technology Evaluation Center (WTEC).
- [9] Chae, J., Park, S. S., Freiheit, T. 2006. Investigation of micro-cutting operations. *International Journal of Machine Tools and Manufacture* 46: 313–332.
- [10] Miao, J. C, Chen, G. L, Lai, X.M, Li, H. T, Li, C. F. 2005. Review of dynamic issues in micro-end-milling' *International Journal of Advanced Manufacturing Technology*, DOI: 10, 1007/s00170-005-0276-6.
- [11] Hsu, T-R, 2002. MEMS & Microsystems- design and Manufacture' McGraw Hill, Singapore.
- [12] Zaman, M. T, Kumar, A. S, Rahman M, Sreeram S. 2006. A three dimensional analytical cutting force model for micro end milling operation. *International Journal of Machine Tools and Manufacture* 46:353–366.
- [13] Rahman, M., Kuma, A. S, Prakash, J. R. S. 2001. Micro-milling of pure copper. *Journal of Material Processing Technology* 116: 39-49.
- [14] Miyaguchi, T., Masuda, M., Takeoka E., Iwabe, H. 2001. Effect of tool stiffness upon tool wear in high spindle speed milling using small ball end mill. *Precision Engineering* 25(2): 145-154
- [15] Tansel, I, Nedbouyan, A, Trujillo, M, Tansel B. 1998. Micro-end-milling-II: Extending tool life with a Smart Workpiece Holder (SWH). *International Journal of Machine Tools and Manufacture* 38: 1437-1448.
- [16] Cook, N. H, Subramanian K, Basile, S. A. 1975. Survey of the state of the art of tool wear sensing techniques. Materials Processing Laboratory, MIT.
- [17] Bao, W. Y. and Tansel, I. N. 2000. Modeling micro-end-milling operations. Part III: influence of tool wear *International Journal of Machine Tool and Manufacture* 40: 2193-2211.
- [18] Bao, W. Y and Tansel, I. N. 2000. Modeling micro-end-milling operations. Part I: analytical cutting force model. *International Journal of Machine Tool and Manufacture* 40: 2155-2173.
- [19] Bao, W. Y. and Tansel, I. N. 2000. Modeling micro-end-milling operations. Part II: tool run-out. *International Journal of Machine Tool and Manufacture* 40: 2175-2192.
- [20] Yuan, Z. J, Zhou M, and Dong, S. 1996. Effect of diamond tool sharpness on minimum cutting thickness and cutting surface integrity in ultraprecision machining' *Journal of Material Processing Technology*, 62 (4):327-330.
- [21] Shimanda, S, Ikawa N, Tanaka H, Ohmuri G, Uchikoshi J., and Yoshinaga H. 1993 Feasibility study on ultimate accuracy in micro-cutting using molecular dynamics simulation, *Annals CIRP* 42 (1):91-94.
- [22] Weule, H., Huntrup, V., and Tritschler, H. 2001. Micro-cutting of steel to meet new requirements in miniaturization, *Annals CIRP* 50: 61-64.
- [23] Lohe, D, Peishl, A, Schilze, V., and Trischler, H. 2000. Micro cutting of steels- interaction of materials properties and process parameters. In *Proceedings of the MicroTEC Conference*, Germany 55:195-204.
- [24] Katayama, S., Toda, M., and Hashimura, M. 1997. Growth model of built-up-edge based on inhomogenities in micro-structure of steel. *International Journal of Japanese Society of Precision Engineering* 31:15-20.
- [25] Zhu, H., and Zhao, J. 1983. Micro manufacturing technology' Chinese Science Publishing Company.
- [26] Lee, K., and Dornfeld, D. A. 2004. A Study of Surface Roughness in the Micro-End-Milling Process, University of California, Berkley
- [27] Alauddin, M., Baradie, M., Hashmi, M. S. J. 1995. Computer-aided analysis of a surface roughness model for end milling,' *Journal of Materials Processing Technology*, 55:123-127.
- [28] Lee, K., Dornfeld, D. A., and Kapoor, S. G. 2002. An experimental study on burr formation in micro milling aluminium and copper. *Trans NAMRI/SME* 30:255-262.
- [29] Dornfeld, D. A. 2004. Strategies for preventing and minimizing burr formation, University of California Berkley.
- [30] Kim, C. J., 2004. Mechanisms of chip formation and cutting dynamics in the micro-scale milling process' PhD dissertation, University of Michigan, USA.
- [31] Liu, X., Jun, M. B. G., DeVor R. E., and Kapoor, S. G. 2004. Cutting mechanisms and their influence on dynamic forces, vibrations and stability in micro-end-milling. In *Proceedings of the ASME International Mechanical Engineering Congress*, California, USA.
- [32] Kim, C J, Maoyor, J. R., and Ni, J. 2004. A static model of chip formation in microscale milling. *Journal of Manufacturing Science and Engineering, Trans ASME* 126(4): 710-718.

Development of Symmetrical Double Sided Linear Wobble Plate Compressor for Bus A/C System

Mohd Suffian, Imran Sairaji, Md. Nor Musa

Faculty of Mechanical Engineering

UTM Skudai, Johor. Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia

Tel: +60-7-5534567, Fax: +60-7-5536688, E-mail: mdnor@fkm.utm.my

Abstract

We have developed a new double sided wobble plate compressor series for medium commercial bus air conditioning system for the first time in the industry. The new wobble plate compressor features a compact design and stable operation compare to other compressor concept thus reducing vibration and noise. The horizontal design reduce the overall space require in the engine compartment as well as weight compared to a typical reciprocating system. Three prototypes have been designed, fabricated and tested in the thermolab of FKM, UTM Skudai. It shows promising result although that there are still many rooms for improvements to be done.

Keyword: Compressor, wobble-plate, symmetrical, air conditioning,

1. Introduction

The highly competitive nature of the refrigeration and air-conditioning industry creates an atmosphere in which systems manufacturers must continuously improve the quality and capability of their products in order to maintain or increase their current market shares. Environmental concerns and higher energy costs persuade governments in many parts of the world to more efficient systems. Thus, customers demand systems that cost less to purchase and to operate at lower costs than at the existing market. The challenge faced by compressor manufacturers is to meet this demands and this require aggressive R&D work.

With the evolution of compressor technology beginning from late 19 century, hundreds of manufactures have emerged throughout the world, producing various models to compress various type of gases. Generally, the terms compressor means a machine for compressing gases such as air, refrigerants, and natural gas which further can be exploited bases on their application. This thesis will focus more on compressors applied in air conditioning system in the automotive industry especially utilized in bus air conditioning system.

Currently, local air conditioning system suppliers for bus air conditioning system obtained their compressor from APM (Automotive Parts Malaysia Sdn. Bhd.) which manufactures the product under license from Zexel. The technology is therefore foreign and all development work is carried out at their parent company. Other air conditioning systems are using conventional reciprocating BOCK compressors which are imported from Germany. On the other

hand, wobble plate type compressor produced by SANDEN are installed in a compact but low capacity system used mainly in cars.

Based on the existing design, a new concept is introduced and developed. The concept introduced has better characteristic while maintaining some of the good features of the existing design such as compactness, vibration and noise reduction. Table 1 shows the new compressor specification compared to the existing DKS-32 swash plate compressor.

Table 1 Comparison Data

Model	DKS-32	P410P-310
Displacement	313 cm ³ /rev	310 cm ³ /rev
Cylinders	10	10
Max. Speed	7000 RPM	3000 RPM
Refrigerant	R134a	
Lubricant	ZXL 100PG (DH PS) 500 cm ³	

2. Description of Technology

The new invented compressor is named as Symmetrical Double Sided Linear Wobble Plate Compressor (SDSLWP). The novelty of this invention is that the compression mechanism on both sets of piston-cylinder assembly has each pair of opposing forces cancelled, both during suction and compression creating a dynamically balance machine.

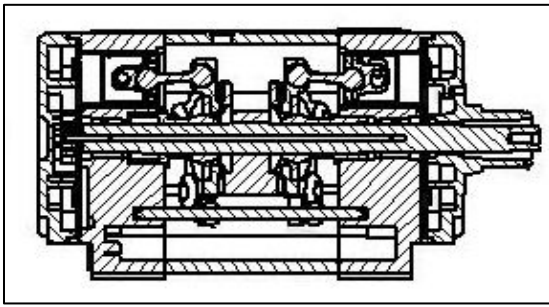


Figure 1 Mechanism of the new invention

Figure 1 shows the mechanism of the new invention. The compressor shaft rotates the rotor producing wobbling motion on the compressor wobble plate which does not rotate due to a bearing mechanism. Next, the wobble plate drives the pistons in reciprocating movement producing suction and discharge effect.

3. Kinematic Analysis

The reciprocating motion of the piston is achieved when the rotation of the shaft is converted to a wobbling motion of the wobble plate. It is an utmost importance for the designer to understand how such motion is achieved. Comprehensive kinematic analysis has been carried out for the mechanism and the reciprocating piston by fellow researcher [1]. Based on the analysis, the tilting angle of the wobble plate and the displacement of pistons relative to shaft rotation can be studied in greater detail. For convenience, stroke position of the nth piston is given by equation 1 with Z equal to stroke position in the z axis, γ is the tilting angle at x-z plane, R is the radius of the connecting rod position at wobble plate, θ is the shaft rotation angle and α is tilting angle at y-z plane..

$$z_s = ((t_w) \cos \gamma + (R_w \sin \theta_i) \sin \gamma) \cos \alpha + (R_w \cos \theta_i) \sin \alpha \quad (1)$$

Figure 2 shows the position of each piston with corresponds to the rotation of shaft.

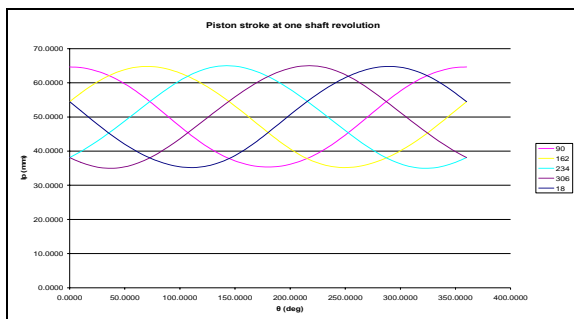


Figure 2 Piston Position at Z-axis

3. Performance

An experimental rig is built to test the performance of the compressor. It comprises of components from existing mini bus air conditioning system rearranged in such a way so

that measurement of important parameters can be made using a data acquisition system. The experimental rig is developed based on British Standard, [2] BS 3122- Part1: 1990-Refrigerant Compressors: Methods of Test for Performance which describes several methods of evaluating a compressor performance. Figure 3 shows the test rig commissioned at thermolab, FKM Skudai.

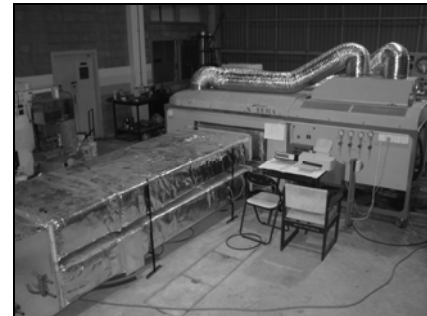


Figure 3 Test Rig

From eight methods available in BS 3122, method F (Flow-meter method) was selected as it was the easiest method to implement without changing the current system. It is also cheaper in terms of cost and easier to set up as the original piping is retained while additional piping is added for the flow meter. However, in the current work, additional features are introduced to the basic design of experimental set-up as in Method F. Pressure of the refrigerant are measured not only at compressor suction and discharge (P1 and P2), but also after the condenser and expansion valve (P3 and P4). Figure 4 shows the actual prototype during experiment.

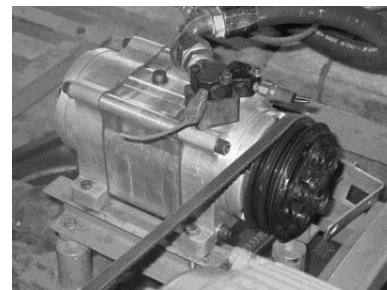


Figure 4 310 cc Compressor

The compressor suction and discharge pressure with respect of time are shown in figure 5. Since the compressor is still in premature state, many improvements have to be made before further performance data can be acquired.

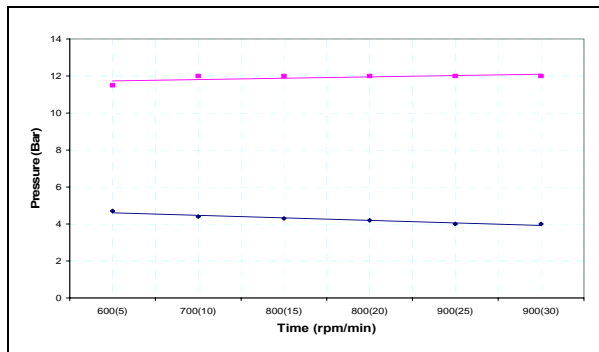


Figure 5 Pressure vs time

4. Conclusion

We have tried to develop a new volumetric compression mechanism called the Symmetrical Double Sided Linear Wobble Plate Compressor as the next generation compressor with several features such as stability and compactness. Several working prototype are also fabricated and under constant review and improvement to further increase the efficiency and durability of the compressor. Compressor design for different capacity and application are also went underway to take advantage of the concept features[3].

5. Acknowledgement

The authors would like to thank the Ministry of Science, Technology and Innovation, Malaysia, for the financial support for this work under the IRPA grant, and to Universiti Teknologi Malaysia for logistical support provided

Reference

- [1] Ahmad, Z.A., Abdul-Latif, A., Syahrom, A., and Musa, M. N. (2003). Study on the Stressing of a New Symmetrical Wobble Plate Compressor: Kinematics and Forces, *Malaysian Science and Technology Congress*. 3-25 September 2003. Kuala Lumpur: COSTAM.
- [2] *Refrigerant Compressors – Part 1 (BS3122-1)* British Standard, UK, 1999.
- [3] Syahrom, A., Musa, M.N., Ahmad, Z.A. and Abdul-Latif, A. (2006). New Symmetrical Wobble-Plate Compressor. International Compressor Engineering Conference at Purdue.

Degradation of phenolic compounds by indigenous isolated microbes

M. Mailin, R. Firdausi

Department of Bioprocess Engineering,
Faculty of Chemical & Natural Resources Engineering,
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor.
e-mail: jh_mailin@yahoo.com

Abstract

The study aims to investigate the performance of isolated phenol-degrading microbes to treat phenol-contaminated wastewaters. From eight types of microbes that were successfully isolated from residential wastewater and oil-contaminated soil, four bacteria showed great potential to be phenol degraders. Four potential candidates were morphologically and biochemically typified, and denoted as RWC-Sma, RWC-Cr1, ISC-Ycr and ISC-Tra. RWC-Cr1 and RWC-Sma were both Gram negative with rod and cocal in shape, respectively; while ISC-Ycr and ISC-Tra were both Gram negative cocci. They are most likely belong to *Pseudomonas* sp, *Alcaligenes* sp, or *Azotobacter* sp, that are reportedly known as excellent phenol degraders. Their ability to be acclimatized in a series of phenol concentration was tested, and RWC-Cr1 which was isolated from residential wastewater demonstrated the highest specific degradation rate, namely 43 m/h phenol, while the rest microbes only showed slight phenol degradation at almost all tested phenol concentrations. Acclimatization of pure culture has enhanced the specific degradation rate from 10% to 75%. Among four isolates, RWC-Cr1 was identified as the best candidate since acclimatization step can be completely skipped in phenol degradation. Therefore, this culture offers a great deal solution to treat phenol and other recalcitrant aromatic compounds in industrial wastewaters.

Keywords: Phenol Degradation, Bacteria, Wastewater, Acclimatization

1. Introduction

During the last few decades, an array of foreign compounds has been introduced into the environment due to the revolution of industrial field. The accumulation of these compounds in environment has resulted in environmental contamination and contributed to many deleterious effects on living systems. Therefore, the removal of toxic compounds from industrial wastewaters is deemed necessary prior to their final discharge for environmental protection. The need to remove these contaminations has led to the development of new technologies that emphasize the detoxification and destruction of pollutants rather than the conventional approach of disposal.

There are varieties of available methods with different removal performance and cost levels to treat wastewaters containing highly toxic organic compounds. Traditionally, wastewaters were treated by costly physiochemical methods, but recently biodegradation has been widely studied and used as a low-cost alternative and offering the possibility of complete mineralization of organic compounds [1]. Most of

these compounds can be used as carbon and energy sources by microorganisms in biological treatment [2]

A wide spectrum of recalcitrant compounds including chlorinated species such as trichloroethylene and certain polychlorinated biphenyls (PCBs), have been shown to be biodegradable under laboratory scale. Currently, biodegradation of aromatic compounds such as phenol, toluene and benzene received severe attention from many industrialists and researchers due to their toxicity and refractory.

Among all those compounds of aromatic, phenol and its derivatives are known as a common constituent of wastewater originating from many industries including pharmaceutical, petroleum and coal refining [3]. The toxicity of phenol has been widely documented and its disastrous effects toward human and environment is greatly concerned. Therefore, this compound was selected as the targeted pollutant in this study.

The objective of this research is to screen microorganisms that have a potential in degrading phenol and also to investigate their performance in phenol biodegradation. This study focused on the effect of acclimatization in the reaction of metabolism.

2. Materials and Methods

2.1 Cultures

Cultures were maintained by periodic subculture on nutrient agar and stored at 4°C. Ramsay Medium which supplemented with NH_4NO_3 , 2.0g/l; KH_2PO_4 , 0.5 g/l; K_2HPO_4 , 1.0 g/l; $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.5 g/l; $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 0.01 g/l; KCl, 0.1 g/l; and Yeast extract, 0.06 g/l, was used as a culture medium [4]. 1 g/l filter-sterilized phenol was aseptically added into the sterile medium prior to inoculation.

2.2 Isolation and Screening of Phenol-degrading Microbes

Samples were taken from residential wastewater treatment plant and industrial oil-contaminated soil. The samples were added into Ramsay medium containing 1 g/l glucose instead of phenol before incubated to allow the enrichment of the culture in the samples. The enriched cultures were further treated with 0.3 g/l phenol. The treatment was carried out at room temperature for 24 hours at 180 rpm. 1 ml of the treated samples were serially diluted and spread evenly over the surface of nutrient agar using sterile L-shape glass spreader.

2.3 Characterization of Isolates

The isolates were identified based on physiological and biochemical characterization. The tests involved namely Gram staining, catalase, oxidase, urease, citrate and indole tests [5], [6] and [7]. Bergey's determinative bacteriology manual was used as the reference to identify the isolates [8].

2.4 Effect of Acclimatization on Phenol Degradation

2.4.1 Phenol Degradation by Acclimatized Culture

Bacteria were inoculated into 100 ml Ramsay Medium containing 0.8 g/l glucose and 0.2 g/l phenol before being incubated at room temperature at 180 rpm. After 24 h, 10 ml of the incubated culture were transferred into a series new sterile medium containing gradually increasing phenol concentrations ranging from 0.2 g/l to 1.0 g/l and gradually

decreasing glucose concentration ranging from 0.8 g/l to 0 g/l.

2.4.2 Phenol Degradation by non-Acclimatized Culture

10 ml of fresh cultures were inoculated into five flasks containing medium supplemented with gradually increasing phenol concentration with gradually decreasing glucose concentration.

2.6 Analytical Method

Samples from fermentation medium were centrifuged at 10,000 rpm for 10 minutes. The supernatant was used for phenol effluent analysis. 2 ml sample was added with 0.3 ml of sodium carbonate (200g/l) and 0.1 ml of Folin-Ciocalteu phenol reagent. After 1 hour incubated at room temperature, the absorbance was measured at 750 nm [9]. Cell density was estimated by optical density measurement at 600 nm using UV- spectrophotometer [10]. Based on the dry weight calibration curve, 1 unit of an optical density is corresponds to approximately 1 g/L of cell.

3. Results and Discussions

3.1 Isolation and Screening of Phenol-degrading Microbes

Residential wastewater and industrial oil-contaminated soils were chosen as the sources of microorganisms because of the tendency of the existence of toxic pollutant. Eight strains of isolate were successfully isolated from both sources of sample after the enrichment stage. The purpose of enrichment stage is to increase the number of microorganisms inside the samples [11].

However, from eight types of microbe, only four strains remained on the medium after further treated with phenol. These four candidates were identified as phenol degraders since the aim of phenol treatment is to ensure only phenol degraders will be isolated in the study [11]. The isolates were denoted as RWC-Sma, RWC-Cr1, ISC-Ycr and ISC-Tra.

3.2 Identification of Isolates

Four potential candidates were morphologically and biochemically typified. As a result, RWC-Cr1 and RWC-

Sma were found to be both Gram negative with rod and cocal in shape, respectively; while ISC-Ycr and ISC-Tra were both Gram negative cocci (Table 1).

According to the biochemical tests results, all isolates were found has the ability to produce oxidase and catalase enzyme but could not produce urease enzyme. Microorganisms that are found to be positive oxidase test might be identified as aerobic or facultative anaerobic bacteria [7]. Meanwhile, oxidase-producing microbes are identified as aerobic bacteria which enable to utilize oxygen as their final electron acceptor in respiratory system. Therefore, all isolates were found to be aerobic microorganisms.

Table 1: Psychological and biochemical characterizations of isolates

TEST	Isolates			
	RWC-Sma	Rwc-Cr1	ISC-Tra	ISC-Ycr
Gram staining	-	-	-	-
Catalase	+	+	+	+
Oxidase	+	+	+	+
Urease	-	-	-	-
Citrate	-	-	+	-
Indole	+/-	+/-	+/-	+

Among four isolates, ISC-Tra exhibited a positive citrate tests which was deducted able to utilize citrate as its source of carbon for growth [7]. However, for indole test, only ISC-Ycr showed a positive result while the rest demonstrated variable test result. Both positive and variable indole test result explained that all isolates are capable to produce indole precursor and have the ability to metabolize ammonium salt as source of nitrogen [7].

Based on physiological and biochemical analyses, the isolates were identified using Bergey’s Manual of Determinative Bacteriology [8]. Accordingly, up to 85 % of the results showed the similarity characteristics with *Pseudomonas* sp, *Alcaligenes* sp or *Azotobacter* sp. As a result, all isolates most likely found to be the genus of the microorganisms. These results are consistent with previous reported studies [10], [12], [13].

3.3 Effect of Acclimatization on Phenol Degradation

3.3.1 Phenol Degradation using Acclimatized Culture

Bacterial was grown in the presence of glucose and adapted to a series of gradually increasing concentrations of phenol with the decreasing of glucose concentrations [14] to allow the microbes to be slowly acclimatized to phenol. The specific degradation rate of isolates is presented in Figure 1.

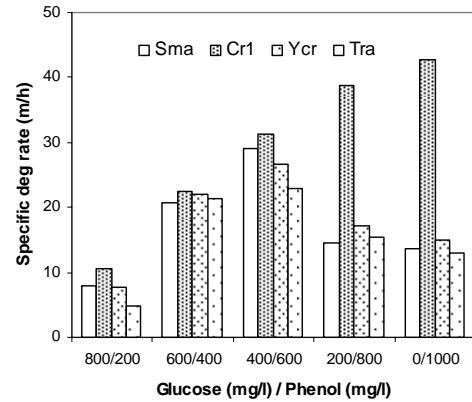


Figure1: Phenol specific degradation rate by acclimatized cultures

The isolates from residential wastewater which designated as RWC-Cr1 demonstrated the increment of specific degradation rate as the concentration of phenol increased. It is likely the higher concentration of phenol would be desirable for RWC-Cr1. These findings suggested that RWC-Cr1 were able to effectively produce more biomass from phenol as sole of carbon source [15]. The highest specific degradation rate was observed at 1000 mg/l phenol, namely 43 m/h. The specific degradation rates are defined as how much concentration of phenol can be degraded within 24 hours by 1 g/l concentration of cell. It is based on the relationship between substrate consumption and cell numbers [15].

Meanwhile, 600 mg/l phenol became the highest phenol concentration level for the rests of isolates to be tolerance to phenol concentration. It was proven by the apparent decrement of specific degradation rate beyond this concentration (Figure1). These results probably explained that high concentrations of phenol have an inhibitory effect on microbial growth [16]. As a result, the metabolism of biodegradation was inhibited resulting in lower phenol removal efficiency [17].

3.3.2 Phenol Degradation using non-Acclimatized Culture

Similar experiment conditions were repeated using non-acclimatized culture to investigate the efficiency of acclimatization in phenol removal. Figure 2 illustrates the

specific degradation rate of isolates using non acclimatized cultures.

Similar to acclimatized phenol biodegradation results, the specific degradation rate of RWC-Cr1 also linearly increased as the concentration of phenol increased. The microbe still can efficiently metabolize phenol to produce more living cells even though non acclimatized cultures were used in the fermentation process. The high number of cells available in the bioreaction contributed to the superiority of RWC-Cr1 in phenol removal [15]. Therefore, acclimatization step can be completely skipped by the culture in phenol biodegradation process since its specific degradation rate in both experiments (acclimatized and non-acclimatized) were nearly equal namely 43 m/h and 38 m/h, respectively.

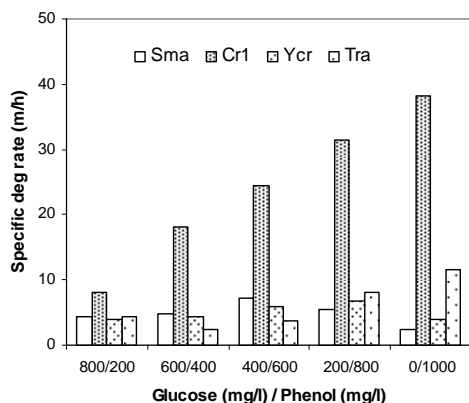


Figure 2: Phenol specific degradation rate by non-acclimatized cultures

However, the rests of microbes ineffectively degraded phenol at all tested phenol concentration. Roughly, the specific degradation rate was just about 5 m/h. Thus, we can conclude here that these microbes entirely require acclimatization step before conducting phenol biodegradation process. In addition, their percentages of phenol reduction were very low particularly beyond to 400 mg/l phenol (data not shown).

3.3.3 Improvement of Acclimatization on Phenol Degradation

The results gained from acclimatized and non acclimatized phenol biodegradation were applied to examine the improvement of acclimatization. All isolates exhibited different improvements of acclimation on phenol degradation at all tested phenol concentration (Figure 3).

ISC-Tra, the isolate from oil-contaminated soil emerged to show the highest improvement, namely 750 % at 400 mg/l phenol concentration. It was obviously proved that

acclimatization improved its degradation ability tremendously. During the acclimatization process, certain enzymes in bacteria are induced so that they are available to take part in the metabolism reaction [18]. Acclimatization also provides sufficient time for small populations of mineralizing organisms to become sufficiently large to bring about a detectable loss of the chemical [19].

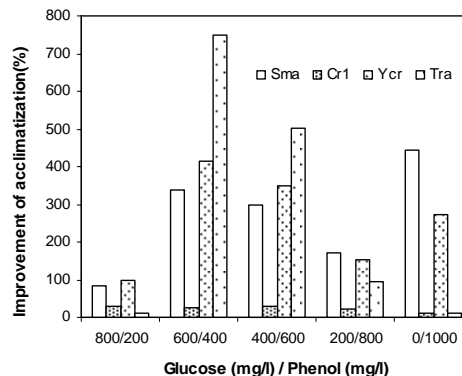


Figure 3: Improvement of acclimatization

The lowest improvement of acclimatization on phenol degradation was exhibited by RWC-Cr1. The range of the improvement values were approximately 12 % - 30 % only. These findings explained that acclimatization insignificantly increased its degradation ability. It can completely consume phenol without conducting such acclimatization before the degradation process. The microbe probably was previously adapted to toxic chemical so that facilitates to the higher efficiency in phenol degradation [20].

4.0 Conclusion

From eight types of microbe that were successfully isolated from residential wastewater and oil-contaminated soil, four bacteria showed a great potential to be a phenol degrader. Based on biochemical and morphological analyses, the isolates are most likely belong to *Pseudomonas* sp, *Alcaligenes* sp, or *Azotobacter* sp. These results are agreed well with the previously reported literature [10], [12] [13].

Generally speaking, acclimatization has improved phenol degradation ability from 10% to 750%. However, RWC-Cr1 emerged as the best candidate to be used in future studies, since acclimatization step can be completely skipped. It is also interesting to note that ISC-Tra demonstrated the highest improvement due to acclimatization.

Acknowledgement

This research was entirely supported by Research Management Centre, UTM, Skudai (vote:75200). We sincerely thank the technical supports provided by Mrs. Siti Zalita, Mr Yaakob, Mr Malik and Mr Mohammad.

References

- [1] Collins, L.D. and Daugulis, A.J. 1997. Characterization and optimization of a two-phase partitioning bioreactor for the biodegradation of phenol. *Applied Microbial Biotechnology*. 48. 18-22.
- [2] Wang, S.J. and Loh, K.C. 1999. Modelling the Role of Metabolic Intermediates in Kinetic of Phenol Biodegradation. *Enzyme and Microbial Technology*. 25. 177-184.
- [3] Nuhoglu, A. and Yalcin, B. 2004. Modelling of phenol removal in a batch reactor. *Process Biochemistry*. 1-7.
- [4] Ramsay, B.A, Cooper, D.G, Margaritis, A. and Zajic, J.E. 1983. "Rhodochorous Bacteria: Biosurfactant Production and Demulsifying Microbial Enhanced Oil Recovery." 61-65.
- [5] Tortora, G.J., Funke, B.R.. and Case, C.L (2004). Microbiology: An introduction. *Pearson Benjamin Cumming*.
- [6] Benson, H.J. 1994. Microbiology applications. 6th edition. W.M.C. Brown Publishers.
- [7] MacFaddin, J.F. 2000. Biochemical test for identification of medical bacteria. Lippincott Williams and Wilkins.
- [8] Goodfellow M. 1994. Bergey's manual of determinative bacteriology 9th edition. *Williams and Wilkins*.
- [9] Box, J.D., 1981. "Investigation of the Folin-Ciocalteu Phenol Reagent for the Determination of Polyphenolic Substances in Natural Waters." *Water Research*. 17.511-524.
- [10] Hannaford, A.M. and Kuek, C. 1999. Aerobic batch degradation of phenol using immobilized *Pseudomonas putida*. *Journal of industrial microbiology and Biotechnology*. 22.121-126.
- [11] Mutzel, A., Reinscheid, U.M., and Antranikian, G. and Muller, R. 1996. Isolation and characterization of a thermophilic bacillus strain, that degraded phenol and cresols as sole carbon source at 70°C. 46. 593-596
- [12] Moustafa El-Sayed Abd El-Hameid Shalaby. 2003. *Biological degradation of substrates mixtures composed of phenol, benzoat, acetate by Burkolderia cepacia G4*. Dissertation Ph.D. University of Carolo-Wilhelmina, Germany.
- [13] Hughes, E. J. and Bayly, R. C. 1983. Control of catechol *meta* cleavage pathway in *Ralstonia eutropha*. *Journal Bacteriol*. 154: 1363-1370.
- [14] Nurdan, K.P., Azmi, T. (2004). Biodegradation of phenol by *Pseudomonas putida immobilized on activated pumice particles*. *Process Biochemistry*. 48. 1807-1814.
- [15] Santos, V.L. and Linardi, V.R.(2003). Biodegradation of phenol by filamentous fungi isolated from industrial effluents-identification and degradation potential. *Process Biochemistry*. 39. 1001-1006.
- [16] Monteiro, A.M.G., Boaventura, A.R. and Rodrigues, A.E. 2000. Phenol biodegradation by *Pseudomonas putida* DSM 548 in a batch reactor. *Biochemistry Engineering Journal*. 6. 45-49.
- [17] Chen, K.C., Lin, Y.H., Chen, W.H. and Liu, Y.C., 2002. Degradation of Phenol by PAA-immobilized *Candida tropicalis*. *Enzyme and Microbial technology*. 31.490-497.
- [18] Kumar, A., Kumar, S. and Kumar, S. (2004) Biodegradation kinetics of phenol and cathecol using *Pseudomonas putida* 1194. *Biochemical Engineering Journal*. 22. 151-159.
- [19] Bruce, A.W., Stephen, H.J. and Martin, A. 1987. Explanation fro the acclimation period preceeding the mineralization of organic chemical in aquatic environments. *Applied and Environmental Microbiology*. 53. 791-796.
- [20] Gonzalez, G., Herrera, G., Garcia, M.T. and Pena, M. 2001. Biodegradation of phenolic industrial wastewater in a fluidized bed bioreactor with immobilized cells of *Pseudomonas putida*. *Bioresearch Technology*. 80. 137-142.

Optimization of Process Conditions for Citric Acid Production Utilizing STP Sludge as Substrate by Liquid State Bioconversion

Md. Zahangir Alam*, Parveen Jamal and Rosnani Saari

Bioenvironmental Engineering Research Unit (BERU), Department of Biotechnology Engineering, Faculty of Engineering, International Islamic University Malaysia (IIUM), Jalan Gombak, 53100 Kuala Lumpur, Malaysia.
Tel: +603-6196457, Fax: +603-61964442, Email: zahangir@iiu.edu.my;

Abstract

Citric acid was produced by *Aspergillus niger* A-SS106 isolated from local sources utilizing sewage treatment plant (STP) sludge as substrate in laboratory scale process. Central composite design (CCD) was carried out to identify the total experiments for optimization of process conditions for citric acid production. The variables involved in the study were temperature, agitation rate, initial pH and inoculum size. A polynomial model was developed using the experimental data for determining the optimum process conditions through response surface methodology with respect to citric acid production. Statistical analysis of results showed the optimized values were: temperature of 30 °C, agitation rate of 350 rpm, initial pH of 4 and inoculum size of 2%. Under the optimum conditions obtained from model equation, the predicted citric acid was 1.51 g/l produced on day 4. The maximum citric acid was 2.9 g/l obtained on day 2 under the optimum conditions.

Keywords

Citric acid, Liquid state bioconversion, Sewage treatment plant sludge, Process optimization

Introduction

The amount of wastewater sludge generated by sewage treatment plants (STP) has been increasing at a rapid pace and contributes the largest organic pollution to water resources as well as to the environment all over the world. In Malaysia, STP sludge contributes about 64.4%, followed by animal husbandry wastes (32.6%), agrobased (1.7%) and industrial effluents (1.3%) in terms of BOD load [1]. Indah Water Konsortium (IWK) treatment plants produce 3.8 million cubic meters of sludge annually as byproduct of waste-water treatment. The management cost is estimated at RM 1 billion [2]. The ever-increasing volume of sludge created every year draws serious attention from society. Therefore a resource recovery process is required to handle the sludge so that the treatment cost will be reduced.

Recently, interest has grown in processes based on resource recovery approach known as recycling and utilization of useful materials from sludge. Sludge contains large amounts of macro and micro-nutrients include calcium (Ca), nitrogen (N), phosphorus (P), potassium (K), calcium (Ca), sodium (Na), and trace metals include manganese (Mn), copper (Cu) and zinc (Zn) [3]. Due to high nutrients contents in sludge and tremendous advancement in fermentation, biotechnology makes its bioconversion into various value added marketable products are possible.

The entire worldwide demand for citric acid is estimated 700,000 tons per annum [4]. Citric acid is widely used in various applications especially in food, pharmaceutical, toiletries and cosmetics industries as

acidulant, stabilizer, flavor enhancer, preservative, antioxidant, emulsifier, and chelating agent [5]. As citric acid is a commodity chemical it is necessary to use inexpensive and readily available materials in industrial processes [6,7]. From this point of view, a large amount of STP sludge with proper nutrients is readily available, cheap (even though negative cost) and renewable materials in Malaysia. It is intended that bioconversion of STP sludge into citric acid would able to solve the problems in terms of disposal and handling of sludge. In this study, sludge which is a by-product of treatment plants becomes a raw material for citric acid production by liquid state bioconversion. Primary benefit of citric acid production from sludge is minimizing the costs of citric acid production and maximizing profits due to negative value of sludge. Therefore, a developed process by statistical optimization was applied to maximize the production and biodegradation of treated sludge using *Aspergillus niger* as the potential citric acid producer fungus.

2. Materials and methods

2.1. Sample Collection and Potential Strain

Raw material, sewage treatment plant (STP) sludge (0.75-1% w/w) was collected from Indah Water Konsortium (IWK) Sewage Treatment Plant, Damansara, Kuala Lumpur, Malaysia. Samples were stored at 4 °C in laboratory cold room for further use. The potential strain *Aspergillus niger* (ASS-106) was obtained from a series experiments of isolation, purification and screening from

previous study [8,9]. The *A. niger* was sub-cultured on PDA plate and incubated at 32°C. Subculture was done once in a month.

2.2. Inoculum Preparation

Aspergillus niger (ASS-106) was obtained from previous study of isolation and screening [10]. Inoculum preparation (spore suspension) was done according to method suggested by Alam, et al [3]. Cultures grown on PDA medium in petri dishes at (30°C±2°C) for 7 days were transferred into Erlenmeyer flask (250 ml) containing 100 mL of sterile distilled water. It was shaken in a rotary shaker at 150 rpm for 24 hours. The suspended fungal cultures were filtered by Whatman #1 filter paper. Finally the filtrate was used as inoculum after measuring the strength (spores/ml) by hemocytometer. All flasks, funnel, filter paper, distilled water were sterilized prior to use.

2.3. Experimental Procedures and Analytical Analysis

For optimization study, a 500 ml of Erlenmeyer flask containing 100 ml of wastewater sludge was used to conduct the lab-scale experiment. The sludge samples as fermentation media with nutrients obtained from previous study were prepared [8,9]. The samples were sterilized at 121°C for 30 min and inoculated with spore suspension of the inoculum of 2% (10⁶cfu/ml). The samples were incubated in the rotary shaker with the process conditions according to the experimental design obtained from statistical software. The sample was harvested once a time to find the interaction among the factors such as temperature, initial pH, inoculum size and agitation for citric acid production.

2.4. Analytical Analysis

The citric acid content in the fermentation broth was determined according to the improved pyridine-acetic anhydride acid method [11], pH was measured using the pH meter at the beginning of the fermentation process. Totals suspended solids was measured according to the standard methods [12]. Chemical oxygen demand (COD), and heavy metals were determined according to the methods described by [13].

2.5. Experimental Design and Statistical Analysis

Experimental design includes statistically design experiments, estimating the coefficients in a mathematical model and predicting the response, and checking the applicability of the model [14]. For this study, experimental design and statistical analysis for process conditions were done by MINITAB (a package from Minitab Inc.). The optimization study was conducted by central composite design (CCD) with 7 central points. Four factors with five levels were considered; temperature, initial pH, agitation rate, and inoculum size. The levels of the variables investigated in this study are shown in Table 1. The data levels for each variables were set based on the investigation of literature for citric acid production from different substrates. A total of 31 sets of experiments (run) were employed in this study to determine the significant factors

affecting the citric acid production are presented in Table 2. The statistical significance of the second-order model equation was determined by the multiple determination coefficient, R²; ANOVA (analysis of variance), p-test and t-test, and response surface methodology.

Table 1. Four factors and five levels of process conditions that will be varied

Factors	Sym-bols	Levels				
		-2	-1	0	1	2
Temp (°C)	X1	20	25	30	35	40
Initial pH	X2	1	2	3	4	5
Agitation rate(rpm)	X3	100	300	400	500	700
Inoculum size (%)	X4	1	2	3	4	5

3. Results and Discussion

Optimization of process conditions for citric acid production from STP sludge by liquid state bioconversion was developed based on values obtained from media optimization from previous study [8,9]. The process conditions were optimized in terms of product formation which is citric acid concentration.

Table 2 represents the results of the optimization study of process conditions. The response of citric acid concentration for optimization study was analyzed on day four. The predicted and experimental citric acid concentration is shown in Table 2. The maximum predicted citric acid was 1.39 g/l (Run 19) obtained on the fourth day of fermentation time. Comparing between the predicted and experimental citric acid yield, the highest experimental citric acid yield was 2.77 g/l (Run 30). However, the predicted citric acid yield was 1.049 g/l (Run 30), which is about 40% lower than the experimental value, was an unbalanced relation in between predicted and experimental values. Therefore, the optimum predicted citric acid 1.388 g/l which was corresponding to the four factors of process conditions was selected for further analysis.

The statistical software MINITAB was used in order to analyze the results shown in Table 2. Quadratic regression equation developed with different factors and their response values considering the linear, quadratic and interaction effects is shown in Equation 1 as follows:

$$Y (\text{Citric acid, g/l}) = -6.33 + 0.188 X_1 + 0.0132 X_2 + 1.51 X_3 + 0.948 X_4 - 0.00388 X_1 X_1 - 0.000072 X_1 X_2 - 0.0133 X_1 X_3 + 0.0203 X_1 X_4 - 0.000016 X_2 X_2 + 0.000500 X_2 X_3 - 0.00120 X_2 X_4 - 0.198 X_3 X_3 - 0.0020 X_3 X_4 - 0.221 X_4 X_4 \quad (1)$$

The model represents a relationship between the dependent variables which is citric acid, Y with the independent variables, namely temperature, agitation rate, initial pH, and inoculum size denoted by X₁, X₂, X₃, X₄. The model was use in order to determine the optimum values of four variables, X₁, X₂, X₃, X₄ based on the highest predicted citric acid concentration.

Table 2. The predicted and experimental citric acid concentration yield

Run order	Temperature (°C)		Agitation (rpm)		Initial pH		Inoculum size (%)		Citric Acid (g/l)	
	Actual	Coded	Actual	Coded	Actual	Coded	Actual	Coded	Experimental	Predicted
1	30	0	250	0	3	0	3	0	1.42	1.27
2	30	0	250	0	3	0	3	0	1.38	1.27
3	30	0	250	0	3	2	3	0	1.06	1.27
4	30	1	250	1	3	-1	3	-1	1.15	1.27
5	30	1	250	1	3	1	3	-1	1.44	1.27
6	30	-1	250	1	3	1	3	1	1.24	1.27
7	30	0	250	0	3	0	3	0	1.31	1.27
8	30	-1	250	-1	3	1	1	-1	0.41	0.53
9	30	-1	250	-1	3	-1	5	1	0.41	0.23
10	30	1	250	-1	5	-1	3	1	0.48	0.56
11	30	2	250	0	1	0	3	0	0.52	0.39
12	35	0	350	0	2	0	2	-2	0.55	0.51
13	35	0	350	-2	4	0	2	0	0.57	0.56
14	35	-1	350	1	4	-1	4	1	0.40	0.37
15	35	-1	350	-1	2	1	4	1	0.43	0.32
16	25	-2	350	0	4	0	4	0	0.35	0.79
17	25	1	350	-1	2	1	4	-1	0.33	0.48
18	25	-1	350	1	2	-1	2	-1	0.39	1.07
19	30	0	350	0	4	0	2	0	1.64	1.39
20	25	0	150	0	4	0	2	0	0.71	0.79
21	25	0	150	0	2	0	4	0	0.58	0.56
22	25	0	150	0	4	0	4	2	0.69	0.67
23	25	1	150	1	2	1	2	1	0.70	0.67
24	35	1	150	1	2	-1	4	1	0.36	0.55
25	35	0	150	0	4	0	2	0	0.32	0.11
26	35	1	150	-1	2	-1	2	-1	0.38	0.25
27	35	1	150	-1	4	1	4	1	0.52	0.40
28	40	0	250	0	3	-2	3	0	0.33	0.46
29	30	-1	100	-1	3	-1	3	-1	0.61	0.77
30	30	-1	400	1	3	1	3	-1	2.77	1.05
31	20	0	250	2	3	0	3	0	1.47	1.30

*significant at 1% and 5% level

The goodness fit of the model was evaluated using determination coefficient, R^2 which was to be 0.90, indicating that 90.0 % of the variability in the response could be explained by the model or only 10.0 % of the total variation were not explained by the model. The closer the value of correlation coefficient, R^2 to 1 expressed better correlation existed between the experimental and predicted value [15]. The value of R is 0.9487 indicating a good agreement existed between the experimental and predicted values of citric acid concentration. The adjusted determination is 81.3% which also indicates high significance of the model [16].

The analysis of variance (ANOVA) was used to confirm the confidence level for this regression model on citric acid production. The confidence level greater than 95% ($P < 0.05$) is highly significant. The results

Table 3. ANOVA for the selected quadratic model

Source	Degree of Freedom	Sum of Squares	Mean Squares	F-value	p>F
Regression	14	4.80277	0.34305	10.30	0.000
Residual error	16	0.53274	0.03330		
Total	30	5.33551			

summarized in Table 3 demonstrates that the quadratic regression models is highly significant indicated by high F-value (10.30) and a very low probability ($p > F = 0.000$).

The t-values, p-values were determined using Minitab software to evaluate the significant level of the coefficients of linear, quadratic and interaction effects on regression model. High value of t-value exemplifies a good interaction of variables and contributes high significance to the model. On the other hand, the low p-value which is less than 0.05 indicates more significant correlation of coefficients [14]. According to Karthikeyan et al.[17], the larger the magnitude of t-value and the smaller the p-value indicate the high significance of the corresponding coefficient. Table 4 presents the t- and p-values on each coefficients of linear, quadratic and interaction of model.

The results revealed that the quadratic effects were significant on the model includes inoculum size, X_4X_4 ($p < 0.000$); initial pH, X_3X_3 ($p < 0.000$); agitation rate, X_2X_2 ($p < 0.008$); and temperature, X_1X_1 ($p < 0.011$). Moreover, the linear terms of initial pH, X_3 ($p < 0.001$); agitation rate, X_2 ($p < 0.007$); and inoculum size, X_4 ($p < 0.026$) gave large effects on the model. The interactive terms with the large effects were temperature-inoculum size, X_1X_4 ($p < 0.041$); and agitation rate-inoculum size, X_2X_4 ($p < 0.018$). The results showed the significant at 1% and 5% confidence levels

Table 4. Interactions of polynomial coefficient corresponding to the p-value and t-value

Predictor	Coefficient	Standard Error Coefficient	t-value	p-value
Constant	-6.334	2.027	-3.12	0.007
Temp, X_1	0.188	0.093	2.01	0.061
Agitation, X_2	0.013	0.004	3.10	0.007*
Initial pH, X_3	1.51	0.39	3.92	0.001*
Inoculum size, X_4	0.95	0.39	2.45	0.026*
X_1X_1	-0.0039	0.0013	-2.86	0.011*
X_1X_2	-0.00007	0.00009	-0.79	0.442
X_1X_3	-0.0133	0.009	-1.46	0.165
X_1X_4	0.0203	0.0091	2.22	0.041*
X_2X_2	-0.00002	0.000005	-3.01	0.008*
X_2X_3	0.0005	0.00046	1.10	0.289
X_2X_4	-0.001	0.000456	-2.64	0.018
X_3X_3	-0.198	0.0339	-5.83	0.000*
X_3X_4	-0.002	0.045	-0.04	0.965
X_4X_4	-0.22	0.034	-6.51	0.000*

From the result discussed, the optimum process conditions were: temperature of 30°C, agitation rate of 350 rpm, initial pH of 4 and inoculum size of 2% for maximum production of citric acid from sludge as major substrate.

4. Conclusions

The results indicated that the statistical optimization of process conditions was able to increase the citric acid

production utilizing STP sludge as the major substrate by liquid state bioconversion. The central composite design (CCD) and regression analysis were able to determine the optimum conditions based on the maximum citric acid production at temperature of 20 °C, agitation rate of 350 rpm, initial pH of 4 and 2.0% of inoculum size. The maximum experimental value of citric acid concentration obtained from the developed process conditions is 2.89 g/l on second day of fermentation period. The organic contains and trace elements present in substrate used is significantly decreases along the fermentation periods implies a potential used of STP sludge as a new carbon source for citric acid production.

References

- DOE, (1996), *Environmental Quality Reports*. Department of Environment, Ministry of Science, Technology and Environmental, Kuala Lumpur, Malaysia, 1996.
- Alam, M. Z., Fakhru'l-Razi, A., Molla, A.H, and Roychoudhury, P. K. 2001. Treatment of Wastewater Sludge by Liquid State Bioconversion Process. *J. Environ. Sci. Health A36(7)*: 1237-1243.
- Alam, M. Z. and Fakhru'l Razi, A. 2003. Biosolids accumulation and biodegradation of domestic wastewater treatment plant sludge by developed liquid state bioconversion process using a batch fermenter. *Water Research 37*: 3569- 3578.
- Ali, S., Haq, I., Qadeer, M. A. and Iqbal, J. 2001. Biosynthesis of citric acid by locally domestic wastewater treatment plant sludge by developed liquid state bioconversion process using a batch fermenter. *Water Research, 37*: 3569- 3578.
- Röhr, M., Kubicek C.P. and Kominek, J. 1996. *Citric acid, in Biotechnology*, 6, 308-345, Verlag Chemie (VCH), Weinheim.
- Haq, I., Khurshid, S., Ali, S., Ashraf, H., Qadeer, M.A. and Rajoka, M.I. 2001. Mutation of *Aspergillus niger* for enhanced citric acid production by blackstrap isolated *Aspergillus niger* using sucrose salt media. *Online J. Biol. Sci.* 1: 178- 181.
- Usami, S. and Fukutomi, N. (1977). Citric acid production by solid fermentation method using sugarcane bagasse and concentrated liquor of pineapple waste, *Hokk. kogaku 55*: 45- 50.
- Jamal, P, Alam, M. Z., Salleh M. R. M., and Akib. M M. 2005. Sewage Treatment Plant Sludge: A Source of Potential Microorganism for Citric Acid Production. *American Journal of Applied Sciences*, 2(8): 1236-1239
- Alam, M. Z., Jamal, P, and Akib. M M. 2005. Optimization of fermentation media for citric acid production using wastewater sludge by surface response methodology. *International Science Congress (ISC) 2005: Engineering and Environmental Sciences (EES)*, 3-6 August, Kuala Lumpur, Malaysia.
- Akib, M. M. 2005. Liquid state bioconversion of sewage treatment plant sludge by liquid state

- bioconversion using *Aspergillus niger*. BSc these, Faculty of Engineering, IIUM, Malaysia.
11. Marier, J.R., and Boulet, M. (1958). Direct determination of citric acid in milk with an improved pyridine- acetic anhydride method. *J Dairy Sci*, 41, 1683-92.
 12. APHA. 1989. *Standard Methods for the Examination of Water and Wastewater*. 17th Edition, American Public Health Association, Washington, D.C
 13. HACH, 2002. *Analysis Handbook*, 4th edition, HACH Company, 185, 355, 383, 749.
 14. Techapun C., Charoenrat, T., Watanabe, M., Sasaki, K. and Poosaran, N. (2002). Optimization of thermostable and alkaline-tolerant cellulose-free xylanase production from agricultural waste by thermotolerant *Streptomyces* sp. Ab106, using the central composite experimental design. *Biochemical Engineering Journal* 12: 99-105.
 15. Mohd Khairizal, M., Rosli, M.I., Roshanida, A.R., Noor Aini, A.R., Nik Azmi, M., Osman, H., Suraini, A.A. and Kamarulzaman, K. (2004). Production of cyclodextrine glucanotransferase (CGTase) from alkalophilic *Bacillus* sp. TS1-1: media optimization using experimental design. *Enzyme and Microbial Technology* 35: 467- 473.
 16. Karthikeyan, R. S., Rakshit, S.K., and Badarajan, A. 1996. Optimization of batch fermentation condition for dextran production. *Bioprocess Eng.* 15: 247- 251.
 17. Akhnazaroza, S., and Kefarov, V. (1982). *Experiment optimization in chemistry and chemical engineering*. Moscow. Mir Publishers.

Feed mill from Solid Waste of Flashing Resulted from Leather Processing Industry

Elly Agustiani¹, Soeprijanto¹, Danawati H.P¹

¹ Department of Chemical Engineering

Sepuluh Nopember Institute Technology of Surabaya, Sukolilo, Surabaya, Indonesia

Tel: +62-81-23027006, Fax: +62-031-5935368, E-mail: cmap@sby.centrin.net.id, ellyagustiani@yahoo.com

Abstract

Solid waste of fleshing resulted from Leather Processing Industry at the industries of tannery in Magetan and Sidoarjo, East Java, Hasn't been optimally processed and this is polluting environment therefore. Each 1 ton of raw hide processed will result 70-230 kg of fleshing as solid waste having potential to pollute environment. (Winter, 1994). The Environmental Impact Assesment Corporation 1995, shows that 586 leather processing industries in Indonesia produced 70.994 ton/year, resulted 78-234 kg of flasing as solid waste. According to Industrial Minister's Decree No.134/M/SK 1998 on prevention and solution to pollution resulted from industrial activities towards living lnvironment and Decree 50/MENLH/II/1996 on odor quality, solid waste of fleshing resulted from small industries of leather processing in Magetan and Sidoarjo giving negative impact on environment due to its mal odor, could spread diseases, polluting earth, water, and air, requiring solution. In 2004, PT. Rajawali In Sidoarjo a Leather Processing Industry made a research to make compost from fleshing waste but it still needed about 2 month which would produce again malodors and the result was still not optimal. Protein content in fleshing is about 37% which can be used as feed supplement. In order to assist leather processing industry to process its waste, LPM ITS and East Java Industrial and Trade Departement cooperate on research to make solid waste of fleshing as cattle feed (feedmill) supplemet. Feed content for feed proportion is based on Gold Coin Group requirement using simple methode. This research is expected to solve problem on fleshing waste polluting environment with high protein content will give a positive image to society towards Leather Processing Industries.

Keywords:

Leather processing, fleshing, feedmill

1. Introduction

Liquid, solid and gas wastes from the activities of home industries of tannery tend to produce pollutions. In order to support our governmental programme to keep continuing environmental live, we need trying to treat or reuse the wastes that are one of the alternative prevention. Solid wastes of fleshing produced by the industries of tannery in the home industry in Magetan and Sidoarjo east java, Indonesia have not been yet optimal treated, therefore those wastes may cause the environmental problems.

According to Decision of Industrial Minister no: 134/M/SK 1988 about the prevention of pollution caused by the activities of industries to the environmental live and *Kep. 50/MENLH/II/1996* concerning standard of odor, therefore, we have to treat the solid wastes of fleshing produced by the industries of tannery in Magetan and Sidoarjo providing nagative impacts to the environment; causing the unpleasant odor; causing the human diseases; and causing soil, water, and air pollutions.

In 2004, PT Rajawali company in Sidoarjo, Surabaya conducted the experiments the use of the fleshing wastes to compost, however, this process took place for a long enough time approximately 2 months, they still produced unpleasant

smell and the products were not optimal. For these reasons, they need solving to treat the fleshing solid wastes that are easily conducted and low cost of operations. In order to help the industries of tannery treat their wastes, therefore, Research Centre ITS (Sepuluh Nopember Institute of Technology Surabaya) collaborated research of the use of the fleshing solid wastes for cow feed supplements with the Department of Industry and Trade, East Java, Indonesia.

The purpose of this experiment was to expect providing the positive perception of the people to the activities of home industries of tannery in the area of Magetan and Sidoarjo, especially the odor pollution and may cause proliferation of disease caused by the great amount of untreated solid wastes. With this information, it was hoped that the industries of tannery can solve the problems of fleshing solid wastes, therefore they may not pollute the environment and may improve high economical values of fleshing wastes containing a high protein.

2. Methods

2.1 Chemicals

The chemicals used in these experiments were all supplied from the following companies: Merck Ltd. Poole Dorset;

Oxoid Basingstoke Hampshire; May & Baker Ltd., Dagenham, Essex; Fisher Scientific Loughborough, Leicestershire; Fluka, Gillingham-Dorset BOC, Glasgow. All chemicals were of general purpose reagent (GPR).

2.3 Fleshing Powder Manufacturing

Fleshing process will produce solid waste in the form of flesh which still contain protein and fat to reach requirement as cattle feed (Lily, 1995) Fat content on solid waste must be produced to 9-10%

2.3.1 Step for manufacturing powder fleshing are :

1. Washing of solid waste in fresh water before putting into a drum by adding 200-300% of water, ZA 3-5% and HCL 2% cattle by litte until the pH is between 6-7 (neutral). Then cheek the sulfide by using Pb paper. Pb paper turn to blue if sulfide still exists) so wash again with fresh water.
2. After being washed, solid waste of fleshing is to be heated at 120°C and pressed at about 2 atm in an autoclaf for about 5 minutes (adding digestion). Solid waste of that autoclaf is then put into an oven and dry it before being around at 200 mesh mill to make powder.

The fleshing powder manufacturing diagram is shown in figure 1.

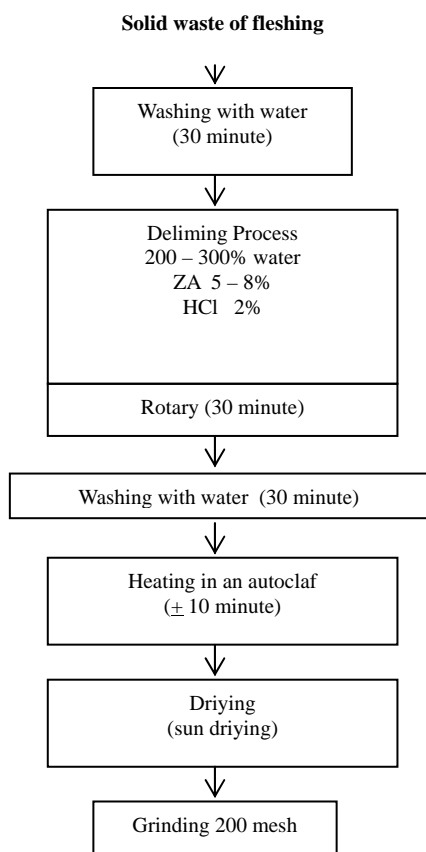


Figure 1. Fleshing Powder Manufacturing Diagram

2.4 Cattle Feed Manufacturing

Manufacturing of cattle feed based on feed proportion using simple method. Feed content for feed proportion is based on Gold Coin Group requirement and use the simple methode.

Banana chip industrial waste which is banana skin can be used as cattle feed. Banana skin containing cellulose and starch is to be added with cattle feed in the form of pellet or powder from fleshing waste containing protein and vitamin at 5 : 1. Supplementing pellet or powder to banana skin is meant to unhealth cattles.

The Composition of cattle’s feed candy, pellet, and powder are shown in Table 1.1 – 1.3

The cattle’s feed in the form of candy, pellet, and powder manufacture diagram is shown in figure 2a, 2b, and 2c.

Basis of the cattle’s feed manufacture is simply methode which the fish powder replace by fleshing powder.

Table 1.1 Cattle’s Feed Candy Composition :

Substance	Composition %			
	Candy I	Candy II	candy III	Candy IV
Corn Seed Powder	45	50	55	60
Fine *Dedak	20	20	20	20
Fleshing Powder	10	10	10	10
Molasses	23	18	13	8
Vitamin	2	2	2	2

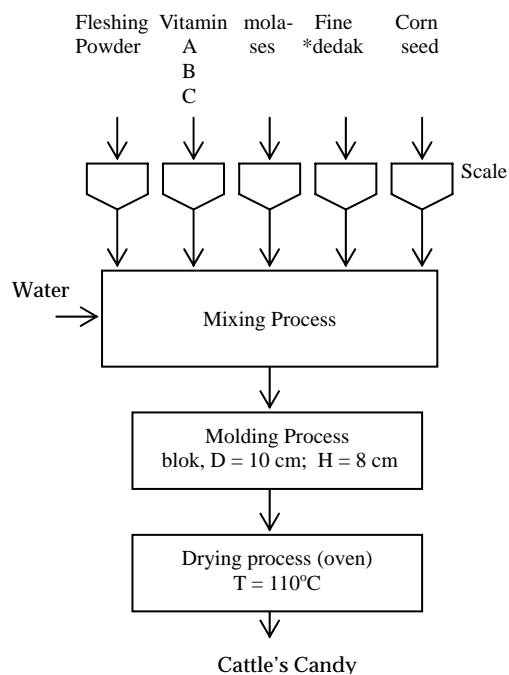


Figure 2a. Cattle’s feed Candy manufacture diagram

Table 1.2 Cattle's Feed Pellet Composition.

Substance	Composition %			
	Pellet I	Pellet II	Pellet III	Pellet IV
Corn Seed Powder	50	55	60	65
Fine *Dedak	20	20	20	20
Fleshing Powder	10	10	10	10
Molasses	18	13	8	3
Vitamin	2	2	2	2

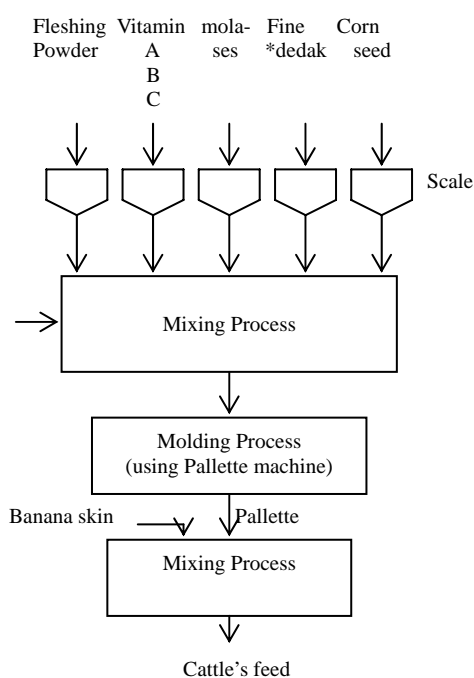


Fig. 2b. Cattle's feed pellet manufacture diagram

Table 1.3 Cattle's Feed powder Composition.

Substance	Composition %			
	I	II	III	IV
Corn Seed Powder	60	65	60	65
Fine *dedak	25	20	28	23
Fleshing Powder	13	13	10	10
Vitamin	2	2	2	2

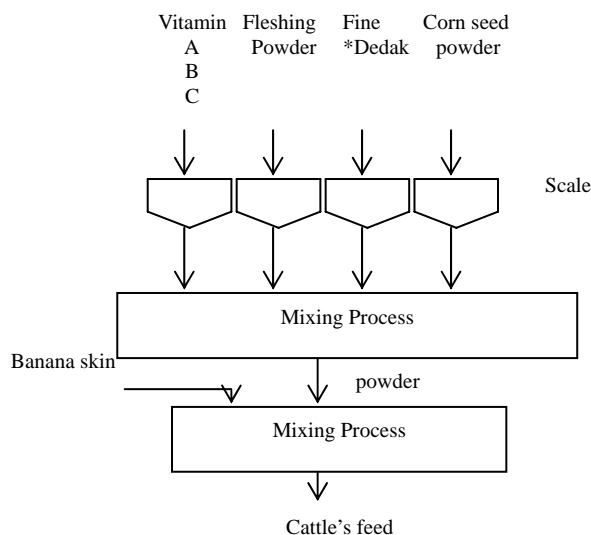


Figure 2c. Cattle's feed powder manufacture diagram

3. Result and Discussion

3.1 Production of Cattle's feed in the form of pellet

It was found that the four varieties of composition of ration in the form of pellet mixed with banana skin were completely consumed by the cows. The ration in the form of pellet could be stored for about 18 months.

3.2 Production of Cattle's feed in the form of candy

It was observed from the four varieties of composition of ration in the form of candy showing that the cows liked the candy very much with the composition I. This ration in the form of candy could be stored approximately 7 months. Tablet cattle feed formulae I (raw sugar yeast content 23%) is the best cattle feed in the tablet form.

3.3 Production of Cattle's feed in the form of powder

The results showed that the four varieties of composition of ration in the form of powder were also completely consumed by the cows. This ration could be stored for approximately 3 months.

Various of cattle's feed expired date can be shown in Figure 3. The best result obtained by Production of Cattle's

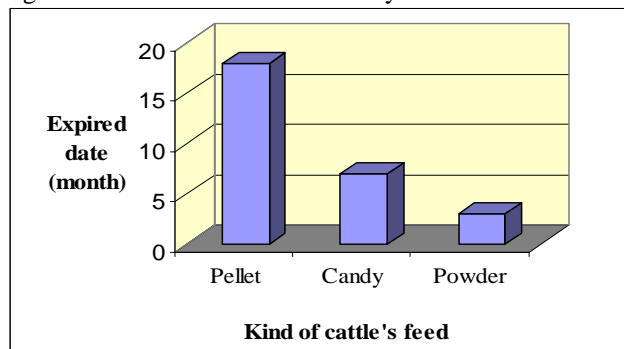


Figure 3 Chart between cattle's feed in the form of pellet, candy, an powder

3. Conclusion

- a. Fleshing waste containing \pm 37% protein is a real economy factor which can be used to make supplement for cattle in the form of powder, pellets and tablets.
- b. Most of the product in the form of powder, pellet, and candy, the cattle like them very much.
- c. Cattle's feed in the form of pellet could be stored the longest amongst those of others approximately 18 months.

Acknowledgments

Thanks to PT Rajawali company in Sidoarjo, Surabaya ;
Research Center ITS, who have supported our work.

References

- [1]----- Balai Penelitian & Pengembangan Industri,
Semarang (1986)
----- Dirjen Industri Aneka Dalam Bapedal (1995)
----- Kep 50/ MENLH/ II/ 1996, (1996). Baku Mutu
Bebauan
----- Kepmen Perindustrian No. 134/ M/ SK (1988).
*Pencegahan dan Penanggulangan Pencemaran
Akibat Kegiatan Usaha Industri Terhadap
Lingkungan Hidup.*
- [2] UNIDO (1997). *Assistance in Establishment and
Operation of a Pilot and Demonstration Plant for
Tannery Effluent Treatment at Estancia Velha R.S.
Brazil*, Project US/ BRA/ 80/ 166/ Contract 85748/ MK.
- [3] Winter, D. (1979). *Techno Economic Study on Measures
to Mitigate the Environmental Impact to the Leather
Industry*, Particularly in Developing Countries, Innsbruck,
Austria.

The Effect of Starch Pretreatment on Cyclodextrins Production

A.M. Mimi Sakinah^{1,2,3}, Rosli Md Illias^{3,*}, A.F. Ismail², Osman Hassan⁴, A.W Zularisam²

¹Faculty of Chemical Engineering and Natural Resources,
University of College Engineering and Technology of Malaysia (KUKTEM),
Kuantan Pahang, Malaysia
Tel: +60-095492322, Fax: +60-095492322, E-mail: mimi@kuktem.edu.my

²Membrane Research Unit, Faculty of Chemical Engineering and Natural Resources,
University Technology of Malaysia, 81310 Skudai Johor, Malaysia
Tel: +60-075535592, Fax: +60-075581463, E-mail: afauzi@utm.my

³Bioprocess Engineering Department, Faculty of Chemical Engineering and Natural Resources,
University Technology of Malaysia, 81310 Skudai Johor, Malaysia
Tel: +60-075495564, Fax: +60-075581463, E-mail: r-rosli@utm.my

⁴School of Chemistry and Food Science, Faculty of Science and Technology,
Universiti Kebangsaan Malaysia, 43000 Bangi Selangor, Malaysia
Tel: +60-0389213805, Fax: +60-0389213232, E-mail: osman@psikukm.my

Abstract

Cyclodextrins (CDs) are inexpensive enzyme-modified starch derivatives which are widely applied in food, cosmetic, pharmaceuticals and plastic industries. This specialty chemical is produced from starch by an enzymatic reaction of cyclodextrin glucosyltransferase (CGTase). Starch consist two macromolecules, which are amylose and amylopectin which are different in chemical and physical nature. Different type of starch sources shows different percentage of amylose and amylopectin contents. Amylose is structurally a long and linear molecule, while amylopectin is molecularly high branched. Normally, the starch was liquefied by jet-cooking before CGTase was added into the vessel. Although the liquefied starch can produce high yield of CDs, there are many disadvantages associated with this method including difficulty in separated CDs from reaction mixture. In order to overcome this problem, the starch pretreatment using moderate heat treated was employed in this study. The production of high concentration of cyclodextrins was obtained by varying the effect of starch pretreatment time in the enzymatic reactor. The optimum starch pretreatment time was observed from the output of cyclodextrins concentration. The high performance liquid chromatography was applied to identify the concentration of cyclodextrins. The experimental result showed that the starch pretreatment was found proportionate with the cyclodextrins yield. This was reason due to the starch transformation which resulted in higher cyclodextrins production as the time increased. Unfortunately, the increasing in time-consuming would increase the energy consumption of the system and cost of the operation.

Keywords: Cyclodextrins; CGTase; Enzymatic reactor; HPLC; Starch pretreatment

1. Introduction

Starch is the major form of carbohydrate storage in plants and is considered as the second largest biomass next to cellulose. The source of starch can be found from tapioca, corn, potato, wheat and rice. Starch is a polysaccharide (meaning "many sugars") of repeating glucose units. The number of glucose molecules joined in a single starch molecule varies from five hundred to several hundred thousand, depending on the type of starch. All starches are

found tightly and basically packed into hydrated granules with about at least one water molecule per glucose unit making up the starch. The shape and the size of the granules depend on the sources of the starch. The major components that make up the starch granules are two macromolecules, amylose and amylopectin. They are different in chemical and physical nature. Amylopectin is responsible for the shape of granule while amylose is mostly found as free molecules. Amylose is structurally a long and linear molecule. Amylose is mainly made of α -(1,4) glycosylic linkage of glucose while amylopectin is highly branched molecule, which has α -(1-4) linked D-glucose chains that are joined to each other by α -(1-6)

*Corresponding Author. E-mail: r-rosli@utm.my
Tel: +60-075495564, Fax: +60-075581463

branch linkages. The α -(1-6) branch linkages constitutes 5% of the linkages in the amylopectin molecule.

Starch reacts with cyclodextrin glycosyltransferase (CGTase) to produce a useful specialty chemical known as cyclodextrins (CDs). Cyclodextrins are cyclic oligosaccharide composed of α -1,4-glycosidic-linked glucosyl residues produced from starch or starch derivatives using cyclodextrin glycosyltransferase (CGTase). CDs can solubilize hydrophobic materials and entrap volatile components by forming inclusion complexes with organic compounds and thus enhance their chemical and physical properties. These properties have led to commercial application of CDs [1-2] in food, pharmaceutical, cosmetic, agricultural and plastic industries as emulsifiers, antioxidants and stabilizing agent [3].

However, the extensive use of CDs is still restricted by high production cost of CDs [4]. Additionally usage of starch as raw material for cyclodextrins production is limited by high viscosity of the medium which hinders stirring and contact between enzyme and substrate. Thus, starch pretreatment by mean of physical, chemical and enzymatic methods is suggested [5, 6]. Parallel to this scenario, this study focused in determining the effect of starch pretreatment on the CDs production by using a pilot scale enzymatic reactor.

2. Materials and Methods

2.1 Production of cyclodextrins

The 10 liters reactor was filled with 4 %wt of raw starch solution mixed with CGTase enzyme. The reaction temperature was conducted under starch gelatinization temperature. This was to avoid starch becoming more viscous and difficult to separate. The temperature and pH of the enzymatic reactor was maintained at 60°C and 6.0, respectively. The experiment was carried out with various starch pretreatment temperature ranging from 60°C to 80°C.

The comparison between starch with pretreatment and starch without pretreatment was also conducted in this research.

2.2 Analysis of CD

The CDs concentrations were determined by using HPLC (Water Assoc.), eluted with acetonitrile: water (70:30) at 1.0 mL/min and detected by refractive index detector (Waters 410). The column temperature was controlled at 30°C. All samples were filtered with Whatman ® nylon membrane filter (0.2 μ m pore size, 13 mm diameter) before injection.

3. Results and Discussion

The starch pretreatment temperature in this study was conducted below its gelatinization temperature in order to obtain the reactive starch granule structure without losing its particle structure. The gelatinization temperature of the tapioca starch was 86°C which was previously determined by using differential scanning calorimetry (DSC). In Figure 1 the amount of cyclodextrins produced from tapioca starch at various temperatures was shown. Based on this figure, the amount of total CDs was continuously increased with reaction time for all condition of experiment. The amount of cyclodextrins was increased from 24% to 96% of starch pretreatment temperature ranging from 60°C to 70°C. However, the starch pretreatment temperature from 70°C to 80°C was observed to substantially increase the CD production from 36% to 111%. In addition, the starch which was heated at 80°C shown the highest amount of cyclodextrins produced compared to the others. This is because, in this region, the temperature was approximately near to the starch gelatinization temperature (above 86°C) where the starch structures was destroyed and produced fully cyclodextrins in high viscosity of medium. On the other hand, only a small amount of cyclodextrins was produced below the 60°C region as the starch structures have not yet change into the reactive structure.

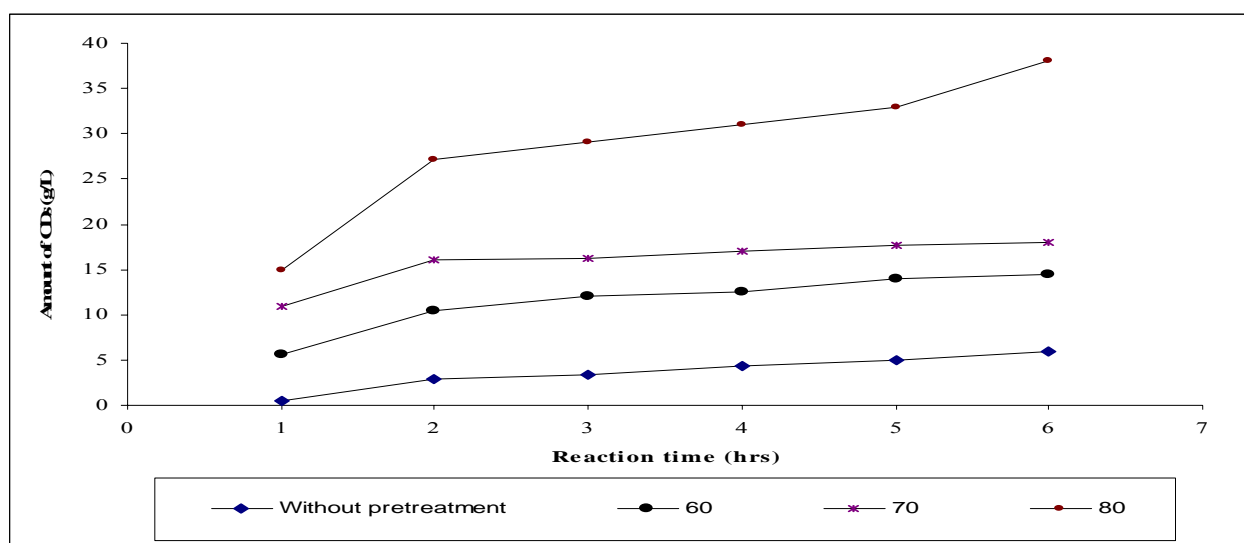


Figure 1. Amount of CDs with different type of starch pretreatment temperature

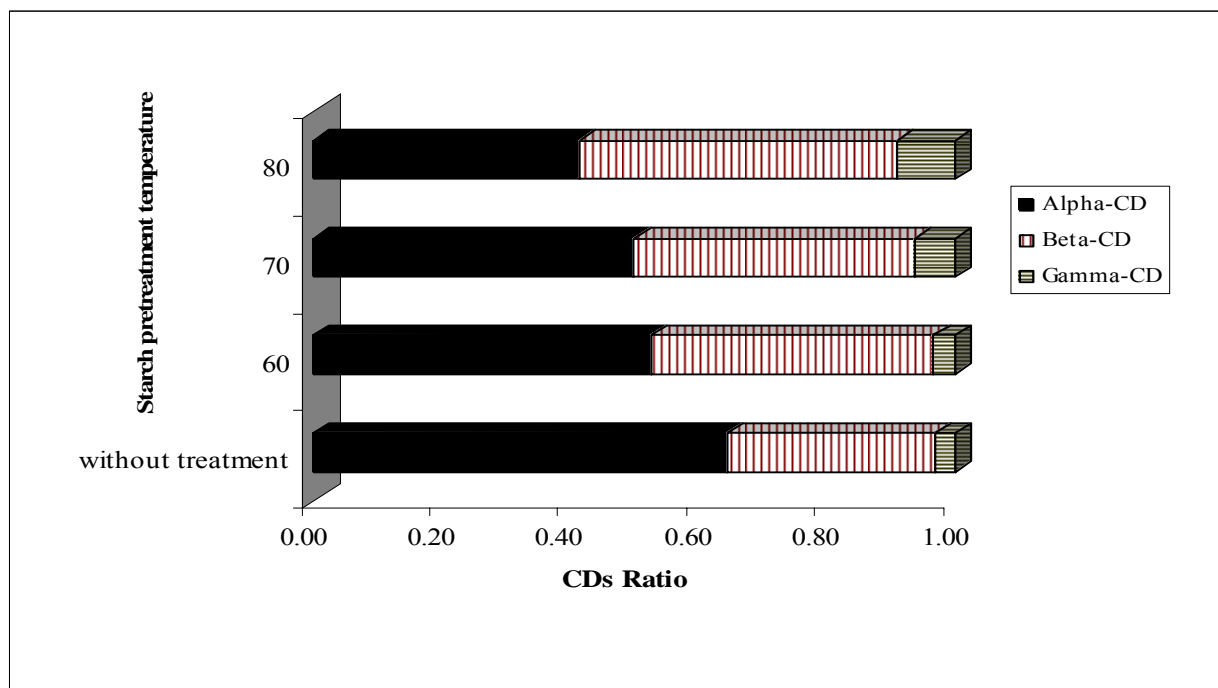


Figure 2. The ratio of CDs with different type of starch pretreatment temperature

Besides that, the comparison between of starch with and without pretreatment method were also investigates in this study. As shown in Figure 1, the starch without pretreatment exhibited the lowest amount of cyclodextrins produced compared to the starch with pretreatment. The similar observation was reported by Kim *et al.* [7] in their study using corn starch.

Even though the amount of cyclodextrins was less produced by non-treated starch, the ratio of α -CD was still higher compared to the starch pretreatment methods. Nevertheless, the increasing in starch pretreatment temperature was not significantly effect the ratio of CDs type (Figure 2). Increasing the starch pretreatment temperature would only cause changes in the ratio of α -CD from 20% to 29%, while the β -CD and γ -CD were 28% to 40% and 50% to 60%, respectively.

Conclusion

The results of these investigations indicated that:

- 1) The optimum starch pretreatment temperature for total amount of CD in this study was 80°C.
- 2) Starch without pretreatment resulted in adverse effect on the total CDs amount due to the molecular structure of the starch granule.
- 3) The ratio of CDs type was not significantly affected by the variation of starch pretreatment temperatures.

References

- [1] Ibrahim, H. M., Yusoff, W. M. W., Hamid, A. A., Illias, R. M., Hassan, O., Omar, O. 2005. Optimization of medium for the production of β -cyclodextrin glucanotransferase using Central Composite Design (CCD). *Process Biochem.* 40:753-758.
- [2] Biwer, A., Antranikian, G., Heinzle, E. 2002. Enzymatic production of cyclodextrins. *Appli. Microbiol Biotechnol.* 59:609-617.
- [3] Szejtli, J. 1997. Utilization of cyclodextrins in industrial products and processes. *Journal of Material Science.* 7(4):575-587
- [4] Kim, T.J., Lee, Y.D., and Kim, H.S. 1993. Enzymatic production of cyclodextrins from milled corn starch in an ultrafiltration membrane bioreactor. *Biotech. and Bioeng.* 41:88-94.
- [5] Biwer, A. & Heinzle, E. 2004. Process modeling and simulation can guide process development: case study α -cyclodextrin. *Enzyme and Microbial Technology.* 34:642-650.
- [6] Charoenlap, N., Dharmsthiti, S., Sirisansaneeyakul, S., Lertsiri, S. 2004. Optimization of cyclodextrin production from sago starch. *Bioresources Technology.* 92: 49-54.
- [7] Kim, T.J., Kim, B.C and Lee, H.S. 1995. Production of cyclodextrins using moderately heat-treated cornstarch. *Enzyme and Microbial Technology.* 17:1057-1061.

Autothermal Reforming of Methane using NiO-CoO/MgO Catalyst

Tutuk D. Kusworo^{1*}, A.R. Songip² and N.A. Saidina Amin²

¹ Department of Chemical Engineering
University of Diponegoro, Semarang, Indonesia
Jl. Prof. Sudharto Tembalang, Semarang Indonesia

² Department of Chemical Engineering
Faculty of Chemical and Natural Resources Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia

Abstract

This paper reports results of studies of the catalytic activity and stability of NiO-CoO/MgO catalyst for autothermal reforming of methane. The catalysts were prepared by an impregnation method and characterized by Temperature Programmed Reduction (TPR), X-ray Diffraction (XRD) and Thermal Gravimetry Analysis (TGA). The TPR and XRD results reveal that the catalyst characteristic is strongly influenced by the Co/Ni ratio. From TPR and TGA analysis, the sintering phenomena did not occur in the autothermal reforming of methane. The results reveal that Co/Ni ratios have a small effect in the catalytic activity for autothermal reforming. Nevertheless, the catalyst showed an optimum performance in this process when its Co/Ni ratio was 0.75.

Keywords: Autothermal reforming, catalyst

1. Introduction

Hydrogen production has received much attention in recent years because hydrogen is considered as a clean energy source (Roh *et al.*, 2002). In the production of hydrogen from methane, highly active and stable catalyst becomes more significant for an on site H₂ generation system or a compact fuel system. Steam reforming of hydrocarbon has been employed frequently in chemical industries for hydrogen production. The disadvantage of steam reforming should be a large extent of heating for the reactor and steam generator. Autothermal reforming (a combination of steam reforming and partial oxidation) has been investigated as potential effective processes to produce hydrogen for fuel cells (Utaka *et al.*, 2003).

By co-feeding steam and oxygen, one can avoid the explosion dangers in partial oxidation, lower the additional steam cost in steam reforming and control the H₂/CO ratio by changing the feeding ratio of steam or oxygen per methane. Many factors such as temperature, CH₄/O₂ ratio, space velocity and CH₄/H₂O ratios are important process variables affecting the reaction of autothermal reforming of methane. It is difficult to find the most important factors and to optimize them. Traditional methods of optimization involved changing one independent variable while fixing the others at a certain level (Larentis *et al.*, 2001).

Therefore, traditional methods have several limitations toward complete optimization, such as time consuming and incapable of reaching the true optimum due especially to interaction among the factors.

In the present paper, the autothermal reforming of methane was investigated experimentally using a NiO-CoO/MgO solid solution catalyst. We studied effects of the Co/Ni ratio, inlet oxygen-to-methane molar ratio and time on stream on the autothermal reforming process.

2. Methods

2.1. Catalyst

The supported NiO-CoO/MgO catalysts were prepared by impregnating the supports Mg (NO₃)₂ · 6 H₂O with aqueous solutions of Ni (NO₃)₂ · 6 H₂O or Co (NO₃)₂ · 6 H₂O followed by overnight drying at 120 °C. The samples were then calcined in a box furnace for four hours at 800 °C. The catalysts were characterized by nitrogen adsorption, XRD, and TPR.

2.2. Apparatus

The partial oxidation of methane to synthesis gas was studied using a fixed bed reactor with 9 mm inside diameter and 300 mm length. The reaction temperature was controlled with a digital controller. The flow of each gas stream was controlled with a mass flow meter (Alicat Scientific). The analysis of effluent gases was performed with a GC-TCD using Porapak N packed column. The reactants used were pure methane (>99.95 %) and oxygen (99.99%). In this study, CH₄ conversion, H₂ selectivity and

* Corresponding Author. E-mail: jokotutuk@yahoo.com,

CO selectivity are defined according to equations (1) - (3).

$$\text{CH}_4\text{conversion} = \frac{\text{mole of methane reacted}}{\text{mole of methane in inlet stream}} \times 100\% \quad (1)$$

$$\text{CO}_{\text{selectivity}} = \frac{\text{mole of carbon atom in CO produced}}{\text{mole of carbon atom reacted from CH}_4} \times 100\% \quad (2)$$

$$\text{H}_2\text{selectivity} = \frac{\text{mole of hydrogen atom in H}_2\text{ produced}}{\text{mole of hydrogen atom reacted from CH}_4} \times 100\% \quad (3)$$

3. Result and Discussion

In this study, the response surface methodology was used as optimization tool. It consists of a group of empirical techniques devoted to the evaluation of relations existing between a cluster of controlled experimental factors and the measured responses according to one or more selected criteria. The maximum values were taken as the response of the design experiments. The optimal process variables of the factors were obtained by solving the regression equation and also by analyzing the response surface contour plot (Liu and Tzeng, 1998 and Sen, 1997).

The TPR curves for the NiO-CoO-MgO catalysts with various Co/Ni ratios are presented in Figure 1. The TPR is quite similar to that observed for a typical NiO-MgO and CoO-MgO complete solid solution from the previous work by Ruckenstein and Hu (1999), Ulla *et al.* (2001) and Wang and Ruckenstein (2002). As expected for the solid solution, the degree of reduction was small. TPR patterns of the NiO-CoO/MgO catalysts with various Co/Ni ratios exhibited three distinct peaks. The maximum peak between 350- 400 °C was attributed to the reduction of relatively free CoO species. The second peak (peak maximum = 480 °C) can be assigned to the free NiO species (Roh *et al.*, 2003). The other peak, with a peak maximum above 800 °C was attributed to the NiO-CoO/MgO solid solution (Choudhary *et al.*, 1997 and Wang and Ruckenstein, 2002).

The NiO-CoO/MgO solid solution reduction occurred at higher temperatures compared to free NiO and free CoO due to the incorporation of Ni²⁺ and Co²⁺ deep in the MgO matrix forming NiO-CoO-MgO solid solution. In addition, several literatures (Tang *et al.*, 1998; Ruckenstein and Hu, 1999 and Wang and Ruckenstein, 2001) also stated that NiO, CoO and MgO are miscible because of similar crystal line structures. They can form ideal solid solution through a mechanism of lattice substitution that leads to a system almost homogeneously mixed at high temperature. Due to the strong interaction between NiO, CoO and MgO, a temperature higher than 800 °C was required to reduce the solid solution because of the irreducibility of MgO and because Ni, Co and Mg shared oxygen in the solid solution.

Figure 1 shows the small TPR peaks between 300 - 500 °C for Co/Ni ratios = 0.3, 0.5, 1 and 1.2, indicating the presence of small amounts of either free NiO or CoO. These

sites may result in sintering and carbon formation during autothermal reforming of methane. The presence of free NiO or CoO suggests that the formation of NiO-CoO-MgO solid solution was not complete for Co/Ni ratios below and above than 0.75. But at Co/Ni ratio = 0.75 only a single TPR peak (above 800 °C) was observed suggesting that the formation of NiO-CoO-MgO solid solution was complete. The formation of NiO-CoO/MgO solid solution made NiO or CoO more intimate with the support, preventing the formation of free NiO or CoO species suspected to be responsible for sintering and carbon formation during methane reforming. The sintering was very harmful for the supported metal catalysts because the aggregation of metal crystallites decreased the number of active site and also accelerated the carbon deposition since large metal ensembles stimulated the carbon deposition.

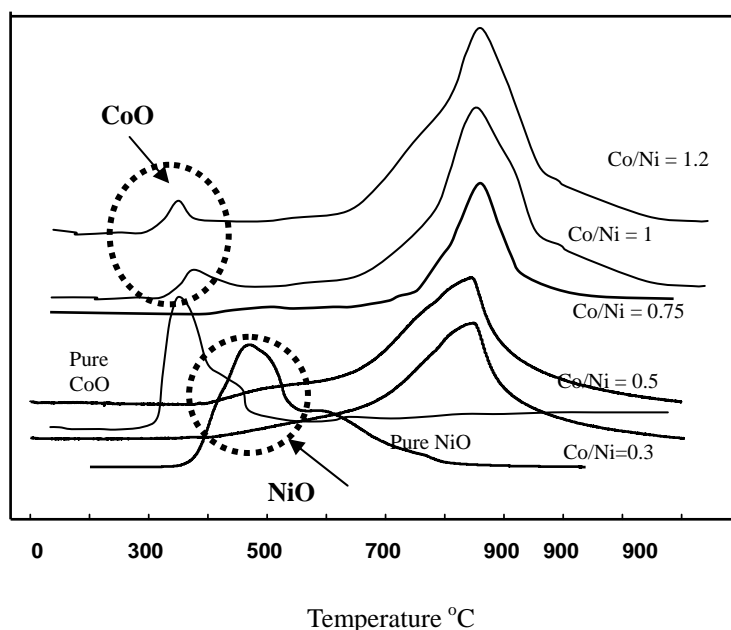


Figure 1: TPR profiles of NiO-CoO-MgO solid solution catalyst with various Co/Ni ratios

The XRD patterns of fresh NiO-CoO/MgO catalyst with Co/Ni ratios and pure MgO are shown in Figure 2. The XRD spectra indicated that all the synthesized NiO-CoO/MgO catalysts exhibited the same pattern of diffraction peaks with the parent MgO catalyst support. There are two possible reasons for this. One is that the impregnation of nickel and cobalt into MgO has not significantly destructed the original structure of the MgO catalysts or the dispersion of nickel and cobalt oxide is too high to be detected using XRD. The other reason is that nickel oxide and cobalt oxide enter the lattice of the MgO to form a uniform solid solution. The XRD result is in agreement with the TPR analysis. The XRD diffractograms can be assigned to NiO, CoO and NiO-CoO/MgO solid solution. The NiO and CoO crystallites did not appear in the catalyst with Co/Ni ratio of 0.75 which implies that the NiO and CoO species were highly dispersed in the lattice of MgO

to form ideal solid solution as similarly reported by Wang and Ruckenstein (2001) and Dong *et al.* (2002). The generations of NiO and CoO crystallites became competitive with Co/Ni ratio above or below 0.75. The fresh catalyst with Co/Ni = 0.3 and 0.5, NiO peaks were identified at $2\theta = 36.44$ and 44.3° (Roh *et al.*, 2002). Meanwhile, in the fresh catalyst with Co/Ni = 1 and 1.2 the peaks shown are identified as CoO peaks at 30.9° and 58.8° (Wang and Ruckenstein, 2002).

Figure 3 shows the results of the TGA experiments as a function of Co/Ni ratio after conducting autothermal reforming of methane. For the ATR reaction, the coking resistant performances of the NiO-CoO/MgO catalysts were more pronounced for all Co/Ni ratios. Autothermal reforming of methane is a combination of partial oxidation and steam reforming, therefore by co-feeding H_2O and O_2 , coke formation can be reduced

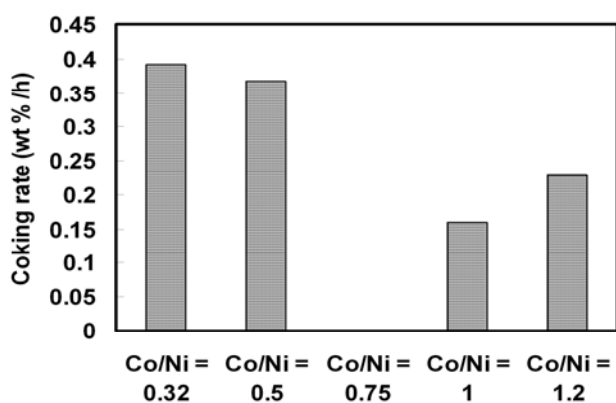


Figure 3: Coke rate on the NiO-CoO/MgO catalyst after conducting autothermal reforming of methane ($T = 850^\circ C$, $CH_4/O_2 = 2.7$, $CH_4/H_2O = 2.7$)

The effects of the Co/Ni ratio of the catalyst are shown in Figure 4. The results (Figure 3) reveal that the Co/Ni ratio has a small effect on the conversion and selectivity. Nevertheless, the catalyst showed an optimum performance in this process when its Co/Ni ratio was 0.75. From TPR and XRD analysis, a complete solid solution of NiO and CoO in MgO was identified in the Co/Ni ratio = 0.75, which is responsible for its superior activity (highest conversion and selectivity). From the TPR and TGA analysis, the sintering phenomena did not occur in the autothermal reforming process. In the autothermal reforming process over the catalyst, the endothermic (steam reforming) and exothermic (partial oxidation) reactions were at least partly coupled by their simultaneous occurrence over the same catalyst. Although the partial oxidation of methane to CO and H_2 was mildly exothermic, a large amount of heat was expected to be produced in a small catalyst volume

because of the high conversion at a very low contact time. However, because of the presence of the steam along with O_2 , the heat that was produced in the exothermic reaction was absorbed in the highly endothermic steam reforming reaction, thus setting a buffer like action that can stabilize the catalyst bed temperature. This is a great practical importance to avoid the hot spot in the reactor. Therefore, this phenomenon can reduce the sintering process and coke formation. Similar observations were also claimed by Choudhary *et al.* (1998a), Ayabe *et al.* (2003), Dong *et al.* (2002) and Supat *et al.* (2003). In addition, Roh *et al.* (2003) also has stated that the autothermal reforming, combination of exothermic (POM) and endothermic (SRM) reactions can hinder the carbon formation in the process.

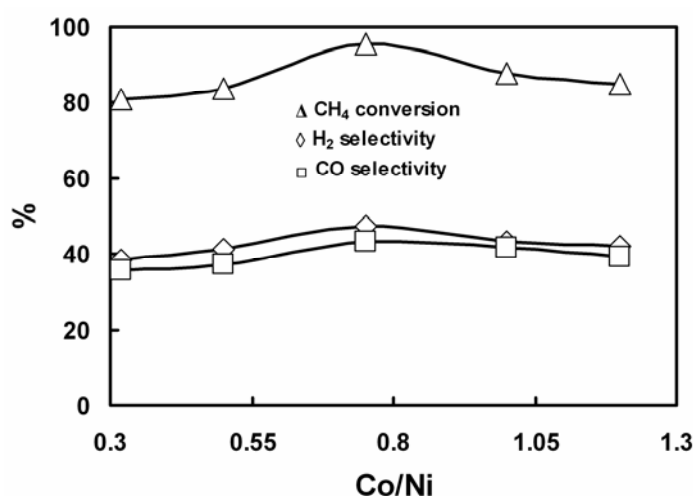


Figure 4: Effect of Co/Ni ratio of the catalyst on the CH_4 conversion, H_2 selectivity and CO selectivity in the autothermal reforming ($GHSV = 5.4 \times 10^4 \text{ cm}^3 \text{ g}^{-1} \text{ h}^{-1}$, $T = 880^\circ C$, $CH_4/O_2 = 2.2$, $CH_4/H_2O = 2.6$)

In order to compare the stability of NiO-CoO/MgO catalyst with commercial catalyst, autothermal reforming of methane was conducted at $880^\circ C$ and $5.4 \times 10^4 \text{ cm}^3 \text{ g}^{-1} \text{ h}^{-1}$ with CH_4/O_2 ratio = 2.2 and CH_4/H_2O ratio = 2.6. Methane conversion, CO and H_2 selectivity with time on stream over both catalysts in autothermal reforming of methane is illustrated in Figure 4. The conversion and selectivity for NiO-CoO/MgO were stable with time on stream. After 12 hours, the methane conversion abruptly dropped to 60 % for the commercial catalyst. It is clear that the commercial catalyst remarkably deactivated with time on stream (Figure 5 (a)) which is most likely due to the carbon formation while NiO-CoO/MgO catalyst maintained its stability. This result indicates that the commercial catalyst was rather more susceptible to carbon formation compared with NiO-CoO/MgO catalyst. Yamazaki *et al.* (1996) and Roh *et al.* (2003) reported a similar result that commercial reforming catalyst deactivated due to carbon formation

under the condition of low steam/carbon ratio ($H_2O/CH_4 = 1$). Therefore, it was confirmed that the NiO-CoO/MgO catalyst was more stable than commercial catalyst. The high stability can be explained as follows. As indicated by XRD and TPR, only solid solution NiO-CoO/MgO species were present in the catalyst with Co/Ni = 0.75. Due to inherent difficulty of reducing a solid solution, the size of the metal clusters generated via the reduction of the solid solution was expectedly small. Being small, these clusters did not stimulate coke formation.

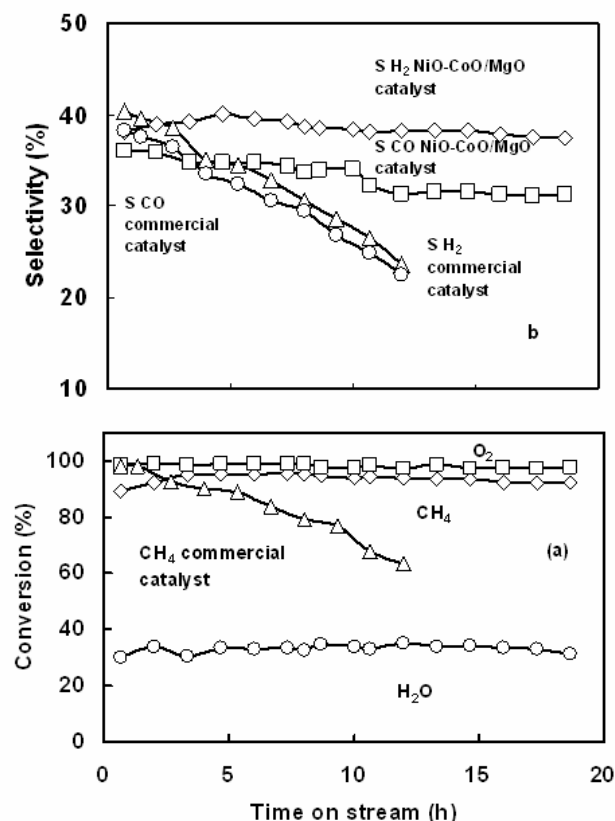


Figure 5: The effects of time on stream on the conversion (a), selectivity (b) in the autothermal reforming of methane at 880 °C for NiO-CoO/MgO and Commercial catalyst (reaction condition: $CH_4/O_2 = 2.2$, $CH_4/H_2O = 2.6$; GHSV = $5.4 \times 10^4 \text{ cm}^3 \text{ g}^{-1} \text{ h}^{-1}$)

4. Conclusion

Studies of the influence of Co/Ni ratios and time on stream on the conversion and selectivity in the autothermal reforming of methane to hydrogen over the NiO-CoO/MgO catalyst lead to the following important conclusions.

1. The formations of complete solid solution of NiO-CoO-MgO catalysts are strongly influenced by the different Co/Ni ratio used.
2. The Co/Ni ratios have a small effect on the conversion and selectivity in the autothermal reforming process.

3. Among the catalyst, NiO-CoO-MgO solid solution catalyst with Co/Ni ratio = 0.75 showed best performance in the autothermal reforming of methane.

References

- [1]. Choudhary, V.R., Rane, V.H. and Rajput, A.M. (1997). Beneficial Effects of Cobalt Addition to Catalysts for Oxidative Conversion of Methane to Syngas. *Applied Catalysis A : General*. 162: 235-238.
- [2]. Choudhary, V.R., Uphade, B.S. and Mamman, A.S. (1998a). Simultaneous Steam and CO₂ Reforming of Methane to Syngas over NiO/MgO/SA-5205 in Presence and Absence of Oxygen. *Applied Catalysis A : General*. 168: 33-46.
- [3]. Choudhary, V.R., Rajput, A. M., Prabhakar, B. and A.S. Mamman, (1998b). Partial Oxidation of Methane to CO and H₂ over Nickel and/ or Cobalt Containing ZrO₂, ThO₂, UO₂, TiO₂ and SiO₂ Catalysts. *Fuel*. 77.(15): 1803-1807.
- [4]. Dong, W.-S., Roh, H.-S., Liu, Z.-W., Jun, K.-W. and Park, S.-E. (2001). Hydrogen Production from Methane Reforming Reactions over Ni/MgO Catalyst. *Bulletin Korean Chemical Society*. 22:1323
- [5]. Dong, W.-S., Roh, H.-S., Jun, K.-W., Park, S.-E. and Oh, Y.-S. (2002). Methane Reforming over Ni/Ce-ZrO₂ Catalysts: Effect of Nickel Content. *Applied Catalysis A : General*. 226: 63-72.
- [6]. Drago, R.S., Jurczyk, K., Kob, N., Bhattacharyya, A. and Masin, J. (1998). Partial Oxidation of Methane to Syngas using NiO-Supported Catalysts. *Catalysis Letters*. 51: 177-181.
- [7]. Roh, H.S., Jun, K.W., Wen, S.D. and Jong, S.C., (2002). Highly Active and Stable Ni/Ce-ZrO₂ Catalyst for H₂ Production from Methane. *Journal of Molecular Catalysis A: Chemical*. 181: 137-142.
- [8]. Roh, H.S., Jun, K.W. and Park, S.-E. (2003). Methane Reforming Reactions over Ni/Ce-ZrO₂/θ-Al₂O₃ Catalysts. *Applied Catalysis A : General*. 251: 275-283.
- [9]. Ruckenstein, E., and Hu, Y.H. (1999). Methane Partial Oxidation over NiO/MgO Solid Solution Catalysts. *Applied Catalysis A : General*. 183: 85-92.
- [10]. Takeguchi, T., Furukawa, S.N, Inoue, M. and Eguchi, K. (2003). Autothermal Reforming of Methane over Ni Catalysts supported over CaO-CeO₂-ZrO₂ Solid Solution. *Applied Catalysis A : General*. 240: 633-642.
- [11]. Tang, S., Lin, J. and Tan, K.L. (1998). Partial Oxidation of Methane to Syngas over Ni/MgO, Ni/CaO, Ni/CeO₂. *Catalysis Letters*. 51: 169-175.
- [12]. Ulla, M.A., Spretz, R., Lombardo, E., Daniell, W. and Knözinger, H. (2001). Catalytic Combustion of Methane on Co/MgO: Characterization of Active Cobalt Sites. *Applied Catalysis B : Environmental*. 29: 217-229.
- [13]. Utaka, T., Al-dress, S.A., Ueda, J., Iwasa, Y., Takeguchi, T., Kikuchi, R. and Eguchi, K. (2003). Partial Oxidation of Methane over Ni Catalysts based on Hexaaluminate-or Perovskite-Type Oxides. *Applied Catalysis A : General*. 247: 125-131.

[14]. Wang, H.Y. and Ruckenstein, E. (2002). Formation of Filamentous Carbon During Methane Decomposition over Co-MgO Catalysts. *Carbon*. 40:1911-1917.

[15]. Supat, K., Chavadej, S., Lobban, L.L. and Mallinson, R.G. (2003). Combined Steam Reforming and Partial Oxidation of Methane to Synthesis Gas under Electrical Discharge. *Industrial Engineering Chemical Resources*. 42: 1654-1661.

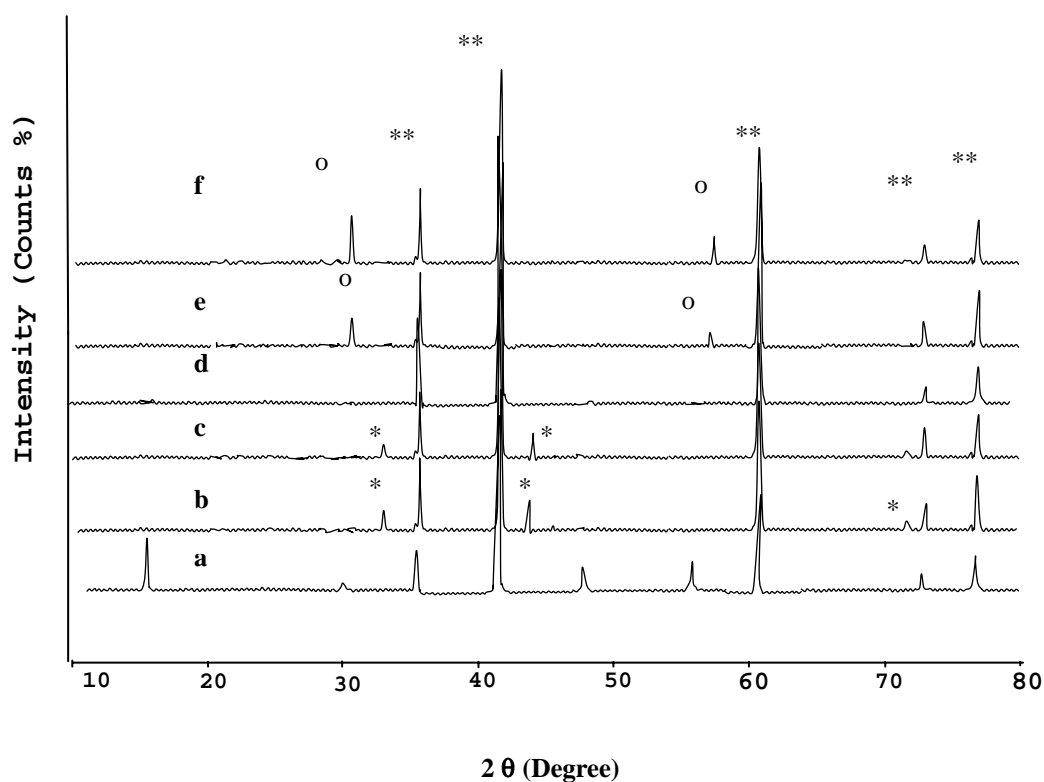


Figure 2: XRD patterns of the NiO-CoO/MgO catalyst: (a) MgO, (b) Co/Ni = 0.3, (c) Co/Ni = 0.5, (d) Co/Ni = 0.75, (e) Co/Ni = 1, (f) Co/Ni = 1.2. (*) represents the diffraction peaks of NiO, (o) represents the diffraction peaks of CoO, (x) represents the diffraction peak of metallic Ni, (**) represents the diffraction peak of the NiO-CoO-MgO solid solution

Statistical Screening of Medium Components for Extracellular Expression of Recombinant Cyclodextrin Glucanotransferase in *Escherichia coli*

Lo Po Kim¹, Tan Cheau Yih¹, Arshad Ahmad², Rosli Md. Illias¹

¹Department of Bioprocess Engineering
Faculty of Chemical and Natural Resources Engineering
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
Tel: 60-7-5535472, Fax: +60-7-5581463, E-mail: lopokim@time.net.my

²Department of Chemical Engineering
Faculty of Chemical and Natural Resources Engineering
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

Abstract

Cyclodextrin glucanotransferase (CGTase) gene from *Bacillus sp. G1 (cgt)* was expressed as extracellular recombinant CGTase in *Escherichia coli*. In order to enhance the production of the recombinant CGTase, a culture medium for the clone *E. coli* harbouring pWH 1520-cgt gene was formulated by statistical approach. This paper basically focused on the medium components screening process for the medium formulation. There are total 15 medium components investigated in the screening process. A 2^{15-10} fractional factorial design (Res. IV) was carried out to identify the significant medium components that will influence the productivity of recombinant CGTase. Finally, only 7 medium components were significant and remained for the medium optimization process. The medium components remains are NZ Amine A, yeast extract, KH_2PO_4 , K_2HPO_4 , glycerol, saccharose and $CuCl_2$. The optimum concentration of these medium components will be further studied in the medium optimization process.

Keyword: recombinant CGTase, extracellular, *E. coli*, medium, screening

1. Introduction

Cyclodextrin glucanotransferase (CGTase; EC 2.4.1.19) is an extracellular enzyme that converts starch, and other polysaccharides to cyclodextrin, CD [1]. Major producers of CGTases belong to *Bacillus sp.* especially aerobic alkalophilic types. The most common available CDs to be synthesized are composed of 6, 7 and 8 glucose units named α -, β - and γ -CDs, respectively [2]. The distribution of α -, β - and γ -CDs is highly dependent on the CGTases used [3].

CDs have a specific torus-shaped structure with hydrophobic interior cavity and hydrophilic exterior surface and therefore can form inclusion complexes with a variety of hydrophobic guest compound. Therefore, CDs can change physical and chemical properties of the encapsulated guest compound [4]. Based on their ability to form inclusion complexes with hydrophobic substances, CDs are widely use in the food, pharmaceutical, cosmetic, toiletries, analytical chemistry and agrochemical industries [5]. For example, CDs can improve the aqueous solubility and stability of various compound, retain the flavour and aroma of compounds in food products, remove unwanted substances from a mixture and also help in controlled-release of drugs [6]. Therefore, CDs became increasingly

important as molecular encapsulator for industrial application.

Given the potential uses of CGTase and their high demand, the need exists for the cloning and recombinant clones that produce CGTase with novel properties and the development of low cost industrial media formulation. In this respect, the fractional factorial design which is used to study the effects of several factors influencing the responses by varying them simultaneously and conducting just a limited numbers of experiments. Screening of the medium components by classical methods, which only involves the change of one variable at a time, is definitely time consuming and the interaction effects between the variables will be neglected. As a result, factorial design was the best choice that will overcome the time consuming difficulty and study the interaction effects between the variables for the screening process. It is well documented that media components greatly influence the production of CGTase in microorganism and their interaction plays an important role in the synthesis of the CGTase [7-9]. Therefore, the objective of this study is to screen the medium component that will enhance the production of recombinant CGTase. The screening study will also be the initial achievement for the medium optimization process in further study.

2. Materials and Methods

2.1 Bacterial strains and growth condition

E. coli K12 N3406 (MoBiTec) that carrying the pJL3-BRP and pWH1520-cgt genes was used throughout this study. In order to analyze the expression level of the CGTase, the cultures were grown at 37°C in modified TB medium (containing 12g/l NZ Amine, 24g/l yeast extract, 8ml/l glycerol, 2.2g/l KH₂PO₄, 9.4g/l K₂HPO₄, 100 µg/ml ampicilin and 34 µg/ml chloramphenicol) as the seed culture. The seed culture was grown overnight at 37°C on a rotary shaker at 200rpm. Then, 1 ml of the seed culture was transferred to the 50ml medium that consisted of various medium components as tabulated at Table 1. The culture was grown at 37°C till OD_{600nm} reached 0.3, and then a final concentration of 1mM xylose and 50µM IPTG were added as the inducers to induce the expression for the recombinant CGTase and BRP, respectively. Finally, the temperature was changed to 20°C and the recombinant CGTase was expressed for 29 hours post-induction time. The expression level of recombinant CGTase was determined by using CGTase assay.

2.2 CGTase Enzyme Assay

Enzyme assay was carried out according to the Kaneko method [10]. The reaction mixture consisted of 1ml of 40mg of soluble starch in 0.1M potassium phosphate buffer (pH 6.0) and 0.1ml of the crude enzyme from the culture, and the mixture was incubated at 60°C for 10 minutes in water bath. The reaction was stopped with 3.5ml of 30mM NaOH. Finally, 0.5ml of 0.02% (w/v) phenoltalein in 5mM Na₂CO₃ was added. After leaving the mixture to stand for 15 minutes at room temperature, the reduction in colour intensity was measured at 550nm by spectrophotometer. A blank lacking the enzyme is tested simultaneously with each batch of samples. 1 unit of enzyme activity was defined as the amount of enzyme that forms 1 µmol of β-CD from soluble starch in one minute.

2.3 Experimental Design and Screening

Experimental design was carried out using Design Expert Software (Stat-Ease Inc., Statistic made easy, Minneapolis, MN, USA, Version 6.0.4). A 2¹⁵⁻¹⁰ fractional factorial design (Res. IV) was utilized to identify the significant medium components that will influence the productivity of extracellular recombinant CGTase from clone *E. coli* K12 N3406. The screening process was designed based on a first order model and there are total 15 medium components investigated in the screening process. The medium components can be categorized into carbon sources (sugar and polyols), nitrogen sources, salts, vitamins and trace elements. A total of 32 sets and 3 sets central point experiments were employed in this study to determine the significant factors affecting the expression level of recombinant CGTase (Table 3). The range of the factors investigated was given in Table 1. The settings of range for factors were based primarily on the investigation of single factors (one at a time) and literature.

3. Results and Discussion

Cyclodextrin glucanotransferase (CGTase) gene from *Bacillus* sp. G1 (*cgt*) was isolated, cloned and expressed in *E. coli* as extracellular recombinant CGTase with the aid of bacteriocin release protein (BRP). In order to enhance the production of recombinant CGTase, a culture medium for *E. coli* harbouring pWH 1520-cgt genes was formulated by statistical approach. Screening process was carried out at the very early stage of the medium formulation by 2¹⁵⁻¹⁰ fractional factorial design (Res. IV) using Design Expert Software. Interpretation of results was analyzed using the analysis of variance (ANOVA) as appropriate to the experimental design used and the effects of medium components studied were initially investigated.

Table 1: Medium components and the concentration level studied in the screening process

Medium components	Label	Unit	Low level (-)	High level (+)	Reference
NZ Amine A	A	g/L	0	24	Modified TB
Yeast Extract	B	g/L	0	48	Modified TB
K ₂ HPO ₄	C	g/L	0	18.8	Modified TB
KH ₂ PO ₄	D	g/L	0	4.4	Modified TB
Glycerol	E	mL/L	0	16	Modified TB
Vitamin B1	F	g/L	0	0.2	[11]
Lactose	G	g/L	0	6	[12]
Saccharose	H	g/L	0	6	Concentration as above
CoCl ₂ ·6H ₂ O	J	mg/L	0	5	[13]
EDTA	K	mg/L	0	10	[14]
ZnSO ₄ ·7H ₂ O	L	mg/L	0	4	[13]
FeSO ₄ ·7H ₂ O	M	g/L	0	0.4	[13]
(NH ₄) ₂ SO ₄	N	g/L	0	8	[11]
CuSO ₄	O	mg/L	0	3	[13]
CuCl ₂	P	mg/L	0	3	Concentration as above

Table 2: Fractional factorial design matrix, measured and predicted enzyme activity

Std	A	B	C	D	E	F	G	H	J	K	L	M	N	O	P	Y _{mea}	Y _{mea} *
	g/l	g/l	g/l	g/l	ml/l	g/l	g/l	g/l	mg/l	mg/l	mg/l	g/l	g/l	mg/l	mg/l	U/ml	U/ml
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12.82
2	24	0	0	0	0	0.2	6	6	0	10	4	0.4	0	0	0	63.42	15.09
3	0	48	0	0	0	0.2	6	0	5	10	0	0	8	3	0	2465.73	51.28
4	24	48	0	0	0	0	0	6	5	0	4	0.4	8	3	0	252.61	20.42
5	0	0	18.8	0	0	0.2	0	6	5	0	4	0	8	0	3	0	12.82
6	24	0	18.8	0	0	0	6	0	5	10	0	0.4	8	0	3	345.89	22.59
7	0	48	18.8	0	0	0	6	6	0	10	4	0	0	3	3	148.35	17.68
8	24	48	18.8	0	0	0.2	0	0	0	0	0	0.4	0	3	3	7720.97	88.80
9	0	0	0	4.4	0	0	6	6	5	0	0	0.4	0	3	3	7.91	13.13
10	24	0	0	4.4	0	0.2	0	0	5	10	4	0	0	3	3	10099.12	101.31
11	0	48	0	4.4	0	0.2	0	6	0	10	0	0.4	8	0	3	7083.62	85.14
12	24	48	0	4.4	0	0	6	0	0	0	4	0	8	0	3	11908.48	109.88
13	0	0	18.8	4.4	0	0.2	6	0	0	0	4	0.4	8	3	0	13821.64	118.26
14	24	0	18.8	4.4	0	0	0	6	0	10	0	0	8	3	0	7416.41	87.07
15	0	48	18.8	4.4	0	0	0	0	5	10	4	0.4	0	0	0	15290.29	124.32
16	24	48	18.8	4.4	0	0.2	6	6	5	0	0	0	0	0	0	16440.80	128.86
17	0	0	0	0	16	0	0	0	0	10	4	0.4	8	3	3	150.27	17.74
18	24	0	0	0	16	0.2	6	6	0	0	0	0	8	3	3	0	12.82
19	0	48	0	0	16	0.2	6	0	5	0	4	0.4	0	0	3	0	12.82
20	24	48	0	0	16	0	0	6	5	10	0	0	0	0	3	2990.52	56.17
21	0	0	18.8	0	16	0.2	0	6	5	10	0	0.4	0	3	3	3.08	12.94
22	24	0	18.8	0	16	0	6	0	5	0	4	0	0	3	3	4541.70	68.60
23	0	48	18.8	0	16	0	6	6	0	0	0	0.4	8	0	0	12467.26	112.39
24	24	48	18.8	0	16	0.2	0	0	0	10	4	0	8	0	0	7764.00	89.04
25	0	0	0	4.4	16	0	6	6	5	10	4	0	8	0	0	43.22	14.41
26	24	0	0	4.4	16	0.2	0	0	5	0	0	0.4	8	0	0	60.68	15.00
27	0	48	0	4.4	16	0.2	0	6	0	0	4	0	0	3	0	2868.43	55.07
28	24	48	0	4.4	16	0	6	0	0	10	0	0.4	0	3	0	376.74	23.26
29	0	0	18.8	4.4	16	0.2	6	0	0	10	0	0	0	0	3	8.38	13.14
30	24	0	18.8	4.4	16	0	0	6	0	0	4	0.4	0	0	3	2326.19	49.91
31	0	48	18.8	4.4	16	0	0	0	5	0	0	0	8	3	3	3125.62	57.36
32	24	48	18.8	4.4	16	0.2	6	6	5	10	4	0.4	8	3	3	4879.84	71.02
33	12	24	9.4	2.2	8	0.1	3	3	2.5	5	2	0.2	4	1.5	1.5	8793.64	94.65
34	12	24	9.4	2.2	8	0.1	3	3	2.5	5	2	0.2	4	1.5	1.5	7436.23	87.18
35	12	24	9.4	2.2	8	0.1	3	3	2.5	5	2	0.2	4	1.5	1.5	7854.43	89.55

*Y (U/ml) calculated using Equation (1).

Table 3: ANOVA for the first order model^a

Source	Sum of squares	Degrees of freedom	Mean squares	F-value	p>F
Model	48474.50	18	2693.03	19.15	< 0.0001
Error	29.11	2	14.56		
Total	54463.10	34			
Curvature	3879.69	1	3879.69	27.60	< 0.0001

^a R, coefficient of correlation = 0.9789

$$\sqrt{Y(U/ml) + 164.41} = 52.85 + 7.14A + 16.12B + 14.33C + 13.85D - 10.24E + 2.37F - 2.52G - 5.04H - 3.91J - 2.71K + 3.30L - 2.67M + 3.23N - 1.80O - 6.45P - 19.06AM - 9.74AN + 10.53AP \quad (1)$$

The details of the ANOVA obtained were shown in Table 3 and P-value (Probability >F) which is less than 0.05 indicated that the model terms are significant. Equation (1) that was only detailed in terms of coded factors is the first order model obtained from the screening process and the calculated enzyme activity was tabulated in Table 2.

Based on Equation (1) that all terms displayed were regardless of their significance in the screening process, the effects of the medium components could be simply identified by the coefficient of the coded model. If the coefficient of model was relatively large, it had more significant effects toward the production of recombinant CGTase. Furthermore, the variable with the positive fitted constant has an enhancer effects towards CGTase production than the one with negative coefficient, which has an inhibitory effects. As a result, from Equation (1) variables A, B, C, D, E, H and P, which had large coefficient (more than 5 when refers to ANOVA which data not shown), were identified as the significant components that will affect the expression level of the recombinant CGTase. Besides, Equation (1) shows that there are significant interaction effects between AM, AN and AP, and these interaction effect will influence the determination of the concentration range that will studied in the optimization process. Based on Table 2, the significant medium components achieved were NZ Amine A (A), Yeast Extract (B), K₂HPO₄ (C), KH₂PO₄ (D),

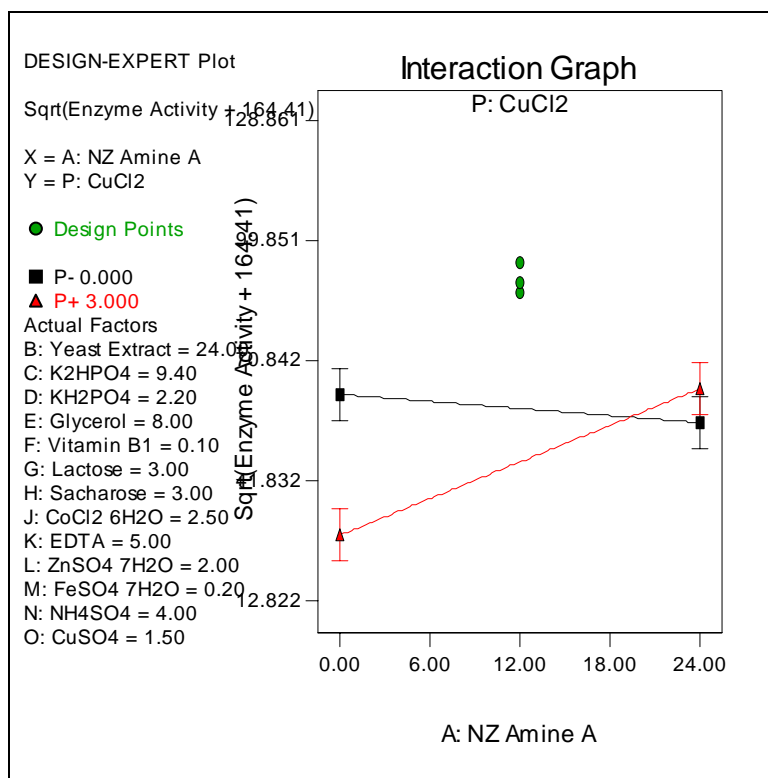
Glycerol (E), Saccharose (H) and CuCl₂ (P). All the medium components existed in the modified TB that was used as an initial medium were found to be significant to influence the production of the recombinant CGTase.

The range of the concentration for the factors studied was clearly consisted of a curvature behavior, which on the other word, the optimum concentration was included in the range, as shown in Table 3. In order to obtain a precise concentration of the medium components screened, a new range of the components that will be studied in the optimization process will be determined based on the results obtained in the screening process. From Equation (1), A, B, C and D had an enhancer effect toward the production of recombinant CGTase when the higher range of the concentration was applied. As a result, the range of the concentration for these component will be set from the middle point till the high level investigated for the optimization process. Contradictory, E, H and P had an inhibitor effect and the range of the concentration will be set from the middle point till the low level involved. Nevertheless, A had some significant interaction effect with P, which means that the range of the concentration for A will depend on P, or the versa vice (Figure 1). Based on Figure 1, when A was set on nearer to the higher range (12 – 24 g/L), P will be set on a higher range (1.5 – 3 mg/L) also. The new range of the concentration determined was tabulated in Table 4.

Table 4: New Range of the Concentration Determined for the Optimization Process

Factor	Components	Unit	Low Range	High Range
A	NZ Amine A	g/L	12	24
B	Yeast Extract	g/L	24	48
C	K ₂ HPO ₄	g/L	9.4	18.8
D	KH ₂ PO ₄	g/L	2.2	4.4
E	Glycerol	mL/L	0	8
H	Sacharose	g/L	0	3
P	CuCl ₂	mg/L	1.5	3

Figure 1: Effect AP (Interaction)



4. Conclusion

In order to formulate a best medium for the clone *E. coli* that will over-express the extracellular recombinant CGTase in a high yield level, the effect of the medium components was investigated in this paper. A medium component screening process was carried out with a 2 level fractional factorial design (Res. IV) was done to determine the significant medium components from the total 15 medium components investigated. Finally, 7 medium components were identified significant and remained for the medium optimization process. The medium components remains are NZ Amine A, yeast extract, KH₂PO₄, K₂HPO₄, glycerol, saccharose and CuCl₂. The optimum concentration of these medium components will be determined in the following medium optimization process.

5. Reference

- [1] Tonkova, A. 1998. Bacterial Cyclodextrin Glucanotransferase. *Enzyme and Microbial Technology* 22:678-686.
- [2] Pedersen, S.; Dijkhuizen, L.; Dijkstra, B. W.; Jersen, B. F. and Jørgensen, S. T. 1995. A Better Enzyme for Cyclodextrins. *CHEMTECH* Dec:19-25.
- [3] Tao, B. Y. 1991. Cyclodextrin Glucanotransferases : Technology and Biocatalyst Design. In American Chemical Society: Enzyme in Biomass Conversion. American Chemical Society, WDC, 372-383.
- [4] Szejtli, J. 1988. *Cyclodextrin Technology*. Netherlands: Kluwer Academic Publishers.
- [5] Biwer, A.; Antranikian, G. and Heinzle, E. 2002. Enzymatic Production of Cyclodextrins. *Applied Microbiology and Biotechnology* 59:609-617.
- [6] Gawande, B.N., Goel, A., Patkar, A.Y. and Nene, S.N. 1999. Purification and Properties of a Novel Raw Starch Degrading Cyclomalto-dextrin Glucanotransferase from *Bacillus firmus*. *Applied Microbiology and Biotechnology* 51:504-509.
- [7] Mahat, M. K.; Illias, R. M.; Rahman, R. A.; Rashid, N. A. A.; Mahmood, N. A. N.; Hassan, O.; Aziz, S. A. and Kamaruddin, K. 2004. Production of Cyclodextrin Glucanotransferase (CGTase) From Alkalophilic *Bacillus* sp. TS1-1: Media Optimization Using Experimental Design. *Enzyme and Microbial Technology* 35:467-473.
- [8] Rahman, A. R.; Illias, R. M.; Nawawi, M. G. M.; Ismail, A. F.; Hassan, O. and Kamaruddin, K. 2004. Optimisation of Growth Medium for the Production of Cyclodextrin Glucanotransferase from *Bacillus stearothermophilus* HR1 Using Response Surface Methodology. *Process Biochemistry* 39:2053-2060.

- [9] Kennedy, M. and Krouse, D. 199). Strategies for Improving Fermentation Medium Performance: A Review. *Journal of Industrial Microbiology & Biotechnology* 23: 456-475.
- [10] Kaneko, T.; Kato, T.; Nakamura, N. and Horikoshi, K. 1987. Spectrophotometric Determination of Cyclozation Activity of β -Cyclodextrin-Forming Cyclodextrin Glucanotransferase. *Journal Japan Society Starch Science* 34:45-48.
- [11] Yuan, H.; Yang, X. and Hua, Z. C. 2004. Optimization of Expression of an Annexin V-Hirudin Chimeric Protein in *Escherichia coli*. *Microbiological Research* 159:147-156.
- [12] Kotik, M.; Kobanová, M.; Marenová, H. and Kyslík, P. 2004. High-level Expression of a Fungal Pyranose Oxidase in High Cell-Density Fed-Batch Cultivations of *Escherichia coli* Using Lactose as Inducer. *Protein Expression and Purification* 36:61-69.
- [13] Levisauskas, D.; Galvanauskas, V.; Henrich, S.; Wilhelm, K.; Volk, N. and Lübbert A. 2003. Model-based Optimization of Viral Capsid Protein Production in Fed-Batch Culture of Recombinant *Escherichia coli*. *Bioprocess and Biosystem Engineering* 25:255-262.
- [14] Riesenberger, D.; Menzel, K.; Schulz, V.; Schumann, K.; Veith, G.; Zuber, G. and Knorre, W. A. 1990. High Cell Density Fermentation of Recombinant *Escherichia coli* Expressing Human Interferon Alpha 1. *Applied Microbiology and Biotechnology* 34:77-82.

Application of Sol-Gel Coating for Solid Phase Microextraction Fiber to the Analysis of Organophosphorus Pesticides

Wan Aini Wan Ibrahim*, Nor Fairolzukry Ahmad Rasdy, Mohd Marsin Sanagi, Ahmedy Abu Naim

Department of Chemistry, Faculty of Science
Universiti Teknologi Malaysia, 81310 UTM, Johor, Malaysia
Tel: +60-7-5535914, Fax: +60-7-5536688, E-mail: fairolzukry@yahoo.com

Abstract

A home-made sol-gel coated solid phase microextraction (SPME) fiber was used for the extraction of selected organophosphorus pesticides (OPPs). The synthesized fibers were characterized by Fourier transform infrared (FTIR), and thermogravimetric analysis (TGA). In this study, a headspace-solid phase microextraction (HS-SPME) mode was used for the analysis of three selected OPPs. The home-made polydimethylsiloxane-poly(vinyl alcohol) (PDMS-PVA) synthesized from sol-gel process compared with commercial polydimethylsiloxane (PDMS) (100 μm thickness). Parameters affecting the HS-SPME performance such as extraction time, extraction temperature, desorption time, desorption temperature and stirring rate were optimized. It was found that HS-SPME using PDMS-PVA fiber gave better extraction of the OPPs studied.

Keywords:

Solid phase microextraction, sol-gel, organophosphorus pesticides

1. Introduction

Organophosphorus compound is one group of pesticides that is largely used in plantations and farms. Organophosphorus pesticides (OPPs) are the most commonly used insecticides, however they are also employed as herbicides and fungicides. Although developed in the early 19th century, it was not until 1932 that the effects of these compounds on insects were discovered [1].

The social concern about the levels of OPPs in the environment and the constant trend observed in current legislations to reduce their maximum residue levels allowed in a variety of matrices is increasing the number of samples to be analyzed well as the need for their accurate determination at very low levels. The help from the advance and improving technology especially in extraction and detection makes the study on OPPs more significance.

Solid phase microextraction (SPME) is an innovative, solvent free technology that is fast, economical and versatile. The introduction of SPME by Pawliszyn [2], create a new dimension for method of extraction and now widely accepted. SPME was introduced to analyze relatively volatile compounds in the

environmental field but now its use has been extended to the analysis of great variety of matrices, such as gas, liquid and solid. Rapid development of this technique resulted in the incorporation of coated fibers into a microsyringe giving rise to the first SPME device [3].

Commercial SPME fibers have generally good extraction properties compared to the other conventional extraction method. However there are some drawbacks and certain aspects of their performance that can be improved. For instance, incomplete desorption that causes carry-over problems and a reduction of the lifetime of the fibers due to direct extraction of higher salt content samples and complex matrices, have been reported in the literature [4]. Furthermore, the low thermal stability, generally bleeding at high temperature ($>250^{\circ}\text{C}$) and solvent stability can lower the extraction and desorption process. The fiber coating lifetime is also limited, usually from about 40 to 100 extractions

Apparently the introduction of sol-gel technology has a promising alternative to overcome those problems. Sol-gel coating technology offers important advantages compared with conventional coating technique; strong adhesion between the coating and the bare fused-silica surface, the porous silica surface structure of sol-gel coating which can provide high surfaces areas and allow for high extraction efficiency and relative ease to

*Corresponding Author. E-mail: wanaini@kimia.fs.utm.my
Tel: +60-7-5535914, Fax: +60-7-5536688

change the composition of the coating which therefore will have different selectivity. This opportunity gives an advantage to improve the performance of SPME fiber. The simplicity of the sol-gel coating procedure provides excellent batch-to-batch reproducibility of the home-made fibers. For optimization, user can vary the coating material mass, altering the precursor or changing the drying time. In this study, a simple and economic methodology for the determination of OPPs using headspace solid phase microextraction (HS-SPME) with a home-made sol-gel polydimethylsiloxane-poly(vinyl alcohol) (PDMS-PVA) fiber coupled to gas chromatography is presented.

2. Experimental

2.1. Chemicals and Reagents

Organophosphorus pesticides (OPPs) standard, chlorpyrifos was purchased from Dr. Ehrenstorfer (Augsburg, Germany), while naled and profenofos were purchased from Sigma-Aldrich (USA). Vinclozolin was used as an internal standard (IS) was purchased from Dr. Ehrenstorfer (Augsburg, Germany). Methanol (HPLC grade) was obtained from Merck (Germany). Stock solutions of 1000 ppm of each OPPs were prepared in methanol. Working standard solutions were prepared by diluting the stock solutions with methanol. The stock solutions and working standard were stored at 4°C when not use. Double-distilled deionized water of at least 18 M Ω was purified by Nano ultra pure water system (Barnstead, USA).

The SPME fiber holder for manual extractions and the fibers of polydimethylsiloxane, 100 μ m film thickness were supplied from Supelco (Bellefonte, USA). The sol-gel coating substances, methyltrimethoxysilane (MTMOS), hydroxyl-terminated polydimethylsiloxane (OH-TPDMS), poly(vinyl alcohol) (PVA), trimethylmethoxysilane (TMMS), poly(methylhydroxysilane) (PMHS), and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (USA).

2.2. Instrumentations

The GC-ECD systems used was a Perkin Elmer XL gas chromatograph (GC) equipped with an electron capture detector (ECD) (San Jose, United State). The capillary column used was a non-polar HP-5 column, 30 m \times 0.32 mm i.d., and 0.25 μ m film thickness. The data has been interpreted using a Perkin Elmer software Turbuchrom Navigator version 4.1

2F12. The column was kept at 180°C for 2 minutes, then ramped at 4°C/min to 200°C, held for 1 minute and finally ramped at 10°C/min to 290 °C, and held for 1 minute. The injection port and detector were set at 280°C and 300°C respectively. Helium as a carrier gas was used at flow rate of 1.0 mL/min and nitrogen as a make-up gas was used at a flow rate of 32.4 mL/min.

Thermogravimetric analysis (TGA) of the PDMS-PVA composite was performed under inert atmosphere (N₂) using a Mettler Toledo Thermal analyzer TG-50 (Greifensee, Switzerland), over the range temperature 50-800°C with a heating rate of 10°C/min. The infrared absorption spectrum of the PDMS-PVA sol gel composite analyzed in KBr pellets between 400 to 4000 cm⁻¹ was obtained using a Perkin Elmer Spectrum One spectrometer (PE-1301F2647, Ueberlingen, Germany).

2.3. SPME Fiber Preparation

Piece of used SPME fiber (~1 cm) was soaked in concentrated sulfuric acid for 3 hours to remove original polymeric layer and expose fused silica core. The fused silica fiber was then rinsed with distilled water and dipped into 1 M NaOH and the fiber was exposed to 0.1 M HCl for 30 minutes. The activated fused silica were then rinsed with distilled water, dried at 120°C and stored for 12 hours in desiccator before use. The PDMS-PVA sol phase was prepared by mixing of MTMOS (600 μ L), OH-TPDMS (300 mg), PVA (80 mg), PMHS (50 μ L), and 95% of aqueous TFA (400 μ L) in a glass tube. These entire materials were being vortex for 2 minutes before the mixture was centrifuged at 12 000 rpm for 5 minutes.

The top sol solution was used for fiber coating. After the treated fiber was dipped vertically into the sol-gel solution for 20 minutes, a sol gel coating was formed on the bare outer surface of the fiber end, about 1 cm. For each fiber, this coating process was repeated three times with a freshly prepared PDMS-PVA sol phase. The coated fibers were end-capped in a 20% (v/v) methanolic solution of trimethylmethoxysilane (TMMS) for 1 minute. Prior to use, the PDMS-PVA fibers were conditioned in the GC injection port at 100°C for 2 hours and then 6 hours at 280°C under 1 mL/min of He.

2.4. HS-SPME Procedure

For HS-SPME mode, sample (2 mL) was placed in a 5 mL sample vial together with a magnetic stirring bar and sealed. The fiber was

introduced carefully directly into the headspace above the aqueous phase and stirred at a rate of 1-point from the total of 10-point maximum speed provided by stirring hotplate, at 20°C for 15 minutes. Then the fiber was retrieved and introduced for a period of 5 minutes into the GC injection port for the analysis. Special comments are given when different conditions were used.

3. Results and Discussion

3.1. Peaks Identification

All the three OPPs studied were eluted within 16 minutes and were well resolved under the temperature program setting. Based on the chromatograms obtained, Naled was eluted first followed by chlorpyrifos and profenofos. The chromatogram of standard solution mixture from HS-SPME is shown in Figure 1.

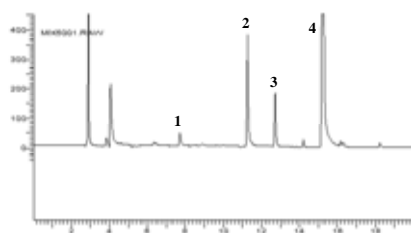


Figure 1: Chromatogram of OPPs studied from HS-SPME. Peaks identification: (1) Naled, (2) IS, (3) Chlorpyrifos and (4) Profenofos

3.2. Characterization

The infrared absorption spectra of the PDMS-PVA sol gel composite is shown in Figure 2 (B) with the corresponding spectra of the conventional PDMS from pure OH-TPDMS (Figure 2A) is also shown for comparison purposes.

The broad absorption band at 3437 cm^{-1} matches the axial stretching of O-H bonds, and is attributed to residual hydroxyl groups not eliminated after the end-capping treatment. This band is also present in the OH-TPDMS with major intensity. Other absorption peaks match both those of sol gel PDMS-PVA and OH-TPDMS: 2964 cm^{-1} corresponds to the stretching of C-H bonds on methyl groups and 1020 , 1197 and 1262 cm^{-1} corresponds to the stretching of Si-O bonds. Peaks at 2168 and 909 cm^{-1} present in sol gel PDMS-PVA spectrum is attributed to Si-H binds from PMHS, used as a surface deactivation reagent.

The thermal stability of the sol-gel PDMS-PVA coating can be assessed from Figure 3, which shows its TGA curve. The second

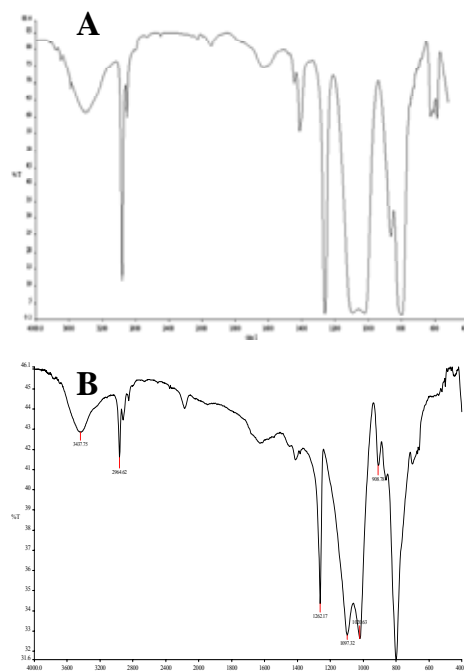


Figure 2: Infrared absorption spectra of: (A) OH-TPDMS and (B) sol gel PDMS-PVA

derivative of this plot was added to the figure to help its interpretation. The first slightly mass loss for PDMS-PVA degradation occurs at 240°C . It is attributed to the release of water and other low molar mass reaction products adsorbed or entrapped inside the pore structure of the polymer not released from the sol-gel during its drying in conventional oven after its preparation. Between 240°C and 330°C , only 2% of the total mass is lost for the sol-gel PDMS-PVA. The main mass loss happens between 400°C and 630°C , corresponding to a loss 50% of the initial mass with no other thermal series visible after this temperature.

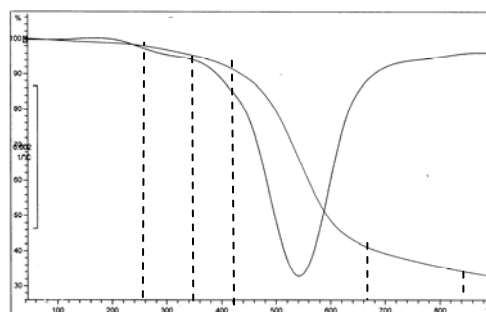


Figure 3: TGA curves (line a) and its second derivative of TGA curves (line b) for the sol-gel PDMS-PVA composite

3.4 HS-SPME Optimization

On the basis of HS-SPME process, a 100 μm commercial PDMS fiber was used as a reference to compare with the sol-gel PDMS-PVA fiber coating. Both were optimized in the same range of conditions. PDMS was chosen in this study because this material has been widely known to have satisfactory extraction for a variety of compounds, especially non-polar one, including OPPs studied [5, 6]. It is hope that the home-made PDMS-PVA fiber from sol-gel process will demonstrate to be good as the commercial PDMS fiber performance if not better. In order to optimize the HS-SPME sampling technique, factors affecting the extraction efficiency such as extraction time, t_{ext} , extraction temperature, T_{ext} , desorption time, t_{des} , desorption temperature, T_{des} , and stirring rate were studied.

The optimum conditions for the extraction of OPPs from water sample were determined for the PDMS-PVA fiber. The optimum extraction condition for the PDMS-PVA fiber as follows: t_{ext} 15 minutes, T_{ext} 50°C, t_{des} 5 minutes, T_{des} 260°C, and stir 1-point. Extraction time and extraction temperature are more dependent on type of stationary phase of the fiber, while desorption time and desorption temperature are more dependent on GC injection port temperature. Stirring was applied to enhance the extraction and speed up extraction time.

3.5 Extraction Efficiency

For a comparison purpose, the optimum HS-SPME conditions for commercial PDMS and sol-gel PDMS-PVA fiber were applied. The peak area obtained for both fibers is shown in Figure 4. It shows that PDMS-PVA fiber can extract more OPPs than PDMS fiber, courtesy of sol-gel coating that provides higher surface area and smaller coating thickness.

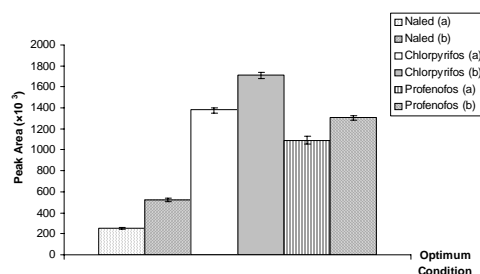


Figure 4: Peak area obtained for both fibers with the optimum condition applied; (a) PDMS and (b) PDMS-PVA

4. Conclusion

Sol-gel chemistry offers a simple and convenient method for coating SPME fiber with desired structure, composition and properties. The presented experimental results clearly demonstrate the potential of sol-gel coating for SPME fibers and chromatography separations for OPPs. The PDMS-PVA fiber showed better extraction of the OPPs than PDMS fiber.

Acknowledgements

We thank Universiti Teknologi Malaysia and The Ministry of Science, Technology and Innovation, Malaysia (MOSTI) for financial supports through IRPA Project 09-02-06-0074 EA211 (Vote No. 74255).

References

- [1] Wagner, S.L. (1989). "The Accute Health Hazzards of Pesticides: Chemistry, Biochemistry, and Toxicology of Pesticides." Corvallis: Oregon State University Cooperative Extension Service
- [2] Arthur, C.L., Pawliszyn, J. (1990). "Solid Phase Microextraction with Thermal Desorption Using Fused Silica Optical Fibers." *Anal Chem.*, 62. 2145-2148
- [3] Pawliszyn, J. (1997). "Solid Phase Microextraction-Theory and Practice." New York: Wiley-VCH
- [4] Bartak, P., Cap, L. (1997). "Determination of Phenols by Solid-Phase Microextraction." *J. Chromatogr. A*, 767. 171
- [5] Goncalves, C., Alpendurada, M.F. (2002). "Multiresidue Method for the Simultaneous Determination of Groups of Pesticides in Ground and Drinking Waters, Using Solid-Phase Microextraction-Gas Chromatography with Electron-Capture and Thermionic Specific Detection." *J. Chromatogr. A*, 968. 177-190
- [6] Lambropoulou, D.A., Albanis, T.A. (2003). "Headspace Solid-Phase Microextraction in Combination with Gas Chromatography-Mass Spectrometry for the Rapid Screening of Organophosphorus Insecticide Residues in Strawberries and Chrries." *J. Chromatogr. A*, 993. 197-203

Electron Microscopes Analysis of Home Grown Multi-Walled Carbon Nanotubes (MWNTs)

Tee Jia Chee¹, Nor Aziah Buang¹, Ahmad Fauzi Ismail^{2,*}

¹Department of Chemistry, Faculty of Science,
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia.
Tel: +60-7-5534123, Fax: +60-7-5534518, E-mail: noraziah@kimia.fs.utm.my

²Membrane Research Unit (MRU), Faculty of Chemical and Natural Resources Engineering,
Universiti Teknologi Malaysia, 81310 UTM, Skudai, Johor, Malaysia.
Tel: +60-7-5535592, Fax: +60-7-5535925, E-mail: afauzi@utm.my

Abstract

A practical and high performance catalytic chemical vapour deposition (CCVD) system was designed and built to produce the multi-walled carbon nanotubes (MWNTs). The electron microscopes are the powerful tools to evaluate the morphology, topology, purity and graphitization of carbon nanotubes (CNTs). Therefore, the objective of this work is to characterize the CNT yields using Scanning Electron Microscope (SEM), Field-Emission Scanning Electron Microscope (FE-SEM) and Transmission Electron Microscope (TEM). The SEM micrographs clearly illustrated the high density growth of CNTs on the substrate, as well as the orientation and the alignment of the nanotubes. The FE-SEM results provided auxiliary view of the surface structure and morphology of the CNTs yield. However, the CNTs yields were defined as MWNTs through the TEM observations. The TEM micrographs showed the degree of graphitization and turbostratic arrangement of the MWNTs. The home grown MWNTs showed high degree of crystallinity and density with configurations of bundles, arrays and coils. The finest MWNT consists of 11 layers of turbostratic graphene wall with inner diameter of 3.57 nm and outer diameter of 11.43 nm as well as distance between layers of 0.33 nm. As a conclusion, high quality and quantity MWNTs were synthesized with the aid of a home built CCVD system.

Keywords:

Multi-Walled Carbon Nanotubes (MWNTs), Scanning Electron Microscope (SEM), Field-Emission Scanning Electron Microscope (FE-SEM) and Transmission Electron Microscope (TEM).

1. Introduction

The past decade has witnessed tremendous effort and progress in the field of carbon nanotubes. Ever since the discovery of carbon nanotubes by Iijima (1991), it has captured the attention of researchers worldwide [1]. Understanding their unique properties and exploring their potential applications have been a main driving force for this area.

Carbon nanotubes (CNTs) are unique tubular structure of nanometer diameter and large length/diameter ratio (~1000). CNTs can be considered conceptually as a prototype one-dimensional (1D) quantum wire [2]. The fundamental building block of carbon nanotubes is the very long all-carbon cylindrical single-walled carbon nanotubes (SWNTs), one atom in wall thickness and tens of atoms around the circumference (typical diameter ~ 1.4 nm).

A single-walled carbon nanotube (SWNT) can be described as a graphene sheet rolled into a seamless hollow cylindrical shape with a high degree of molecular perfection. A multi-walled carbon nanotube (MWNT) consists of concentric cylinders with an interlayer spacing of 0.34 nm and a diameter of typically 10 – 20 nm, and these were the first type of nanotubes observed experimentally. The lengths of two types of nanotubes can be up to several hundreds of microns or even centimeters [3]. They can either occur in individual tubes or bundle up into 'ropes' due to the Van der Waals interaction between their side walls.

The main hindrance to employing carbon nanotubes (CNTs) in real world is the inability to control the growth of the nanotubes and to grow bulk amounts of carbon nanotubes. There are three main techniques to grow carbon nanotubes: arc-discharge, laser ablation and chemical vapour deposition (CVD). The first two methods are high temperature processes that produce high quality CNTs, but they cannot grow mass quantities of nanotubes within a reasonable amount of time. The CVD technique is able to grow bulk amounts of nanotubes and arrays of multi-walled

* Corresponding Author. E-mail:afauzi@utm.my,
Tel: +60-7-5535592, Fax: +60-7-5535925

carbon nanotubes (MWNTs). However, these nanotubes contain a vast amount of defects along the length of the tubes due to the relatively low synthesis temperature of 600 – 1200 °C.

Nevertheless, some progress has been recently obtained, the chemical vapour deposition (CVD) has been modified by applying various supported metals catalysts in the production of CNTs. The catalytic chemical vapour deposition (CCVD) method supplies CNTs in high yield and low costs, but also at controlling the CNTs characteristics and morphologies [4]. Being a catalytic process, the combinations of transition metals and support can be changed depending on the characteristics required, such as the alignment and diameter of the nanotubes. The CCVD synthesis of CNTs can be carried out at low temperature and ambient pressure.

2. Methods

2.1 Synthesis of Carbon Nanotubes (CNTs)

In this work, alumina (Al_2O_3) beads obtained from commercial source were used as catalyst support. Before the addition of catalyst onto the supports, the support was dried to eliminate water vapour and impurities. The catalysts were prepared by impregnation method. The support was immersed with either Co or Fe solution (containing specific wt % of catalyst) until saturation and dried overnight. The prepared supported catalysts were then used in the synthesis of CNTs.

A home-built reactor system has been designed to carry out the synthesis of CNTs. The reactor is a simple and economical thermal catalytic chemical vapour deposition (CCVD) system. It is easy to handle and do not require expensive power supply and high pressure reaction chamber. The experiments to produce CNTs are carried out in a horizontal tube furnace at atmospheric pressure. The supported catalyst was placed inside a quartz tube. Acetylene gas, C_2H_2 was passed through to react with the catalyst. The pyrolysis of C_2H_2 was performed at 700 °C. Prior to sampling, the reaction chamber was cooled down to room temperature. The samples were collected as black powder on the catalyst.

2.2 Electron Microscopes Analysis of CNTs

In the context of CNTs characterization, Scanning Electron Microscopy (SEM) is the most important instrument for the investigation of topology and morphology of the specimens. In this work, a Philips XL 40 SEM was utilized to investigate all of the supported catalysts and as-grown CNTs with 30.0 keV probe under working distance of about 10 mm.

The Field Emission-Scanning Electron Microscope (FE-SEM) provides a view of the surface or topology of the sample. The carbon nanotubes produced in this work were examined by a JOEL JSM-6700F model FE-SEM.

For characterization of CNTs, Transmission Electron Microscopy (TEM) is used for the determination of the type

of carbon nanotubes (SWNT or MWNT), diameter of the tube, amount of graphitization and amorphous deposits and other structural properties of the tubes such as bending and flattening of the tubes. A JOEL JEM-2010 model TEM was used to analyze the as-grown CNTs.

3. Results and Discussion

3.1 Scanning Electron Microscope (SEM) Analysis of As-grown CNTs

From the SEM images, the carbon nanotubes (CNTs) can be observed as are highly graphitized tubules but the carbon nanofibers is amorphous fibrous layers [5]. Therefore, through the SEM analysis, we can differentiate the growth of CNTs from the growth of carbon nanofibers (CNFs) on the surface of supported catalysts. The SEM micrographs demonstrate images of the CNT growth density and the nanotubes texture.

SEM micrographs of as-grown carbon deposits over different Al_2O_3 -Co catalysts are shown in Figure 1. SEM analyses revealed that the density and the nature of CNTs produced were unique in each case. In Figure 1, the growth of CNTs over Al_2O_3 -Co catalyst was clearly observed, the Co particles were not visible in the micrograph as being covered up by the highly dense CNTs. From Figure 1, the CNTs can be seen as bundles of tubes with fluffy and spongy texture. The CNTs possess high quality and density with diameter of approximately 53 nm and length less than 10 microns. Obviously, the as-grown nanotube bundles are pure without purification, as no amorphous carbon and carbonaceous particles deposited on the catalyst. Here, the amorphous carbon was defined as carbonaceous deposit which is not in the form of ordered nanotubes and graphite.

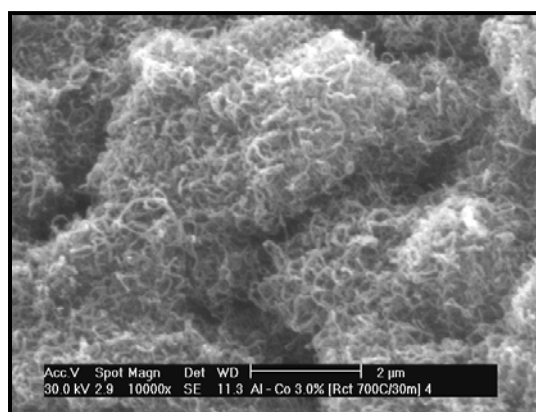


Figure 1. Bundles of as-grown CNTs over Al_2O_3 -Co catalyst (10000 \times).

The Al_2O_3 -Fe catalyst shows very encouraging CNTs growth as depicted by the SEM image (Figure 2). The SEM image conveys the contiguous mass of nanotubes with high uniformity and the morphology and smooth surface of the curved nanotubes. The nanotubes grow extremely long with low diameter distribution, the estimated diameter is 70 nm. The texture of the nanotubes can be seen as fluffy and in bundle form.

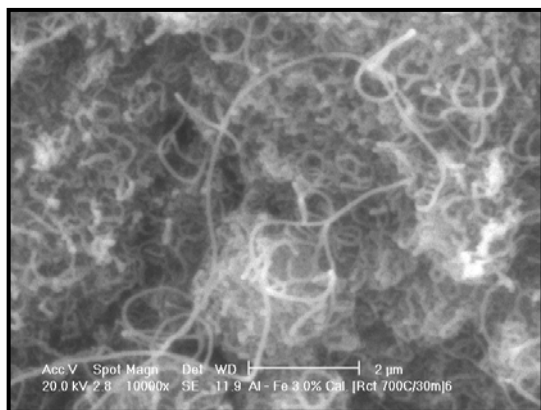


Figure 2. As-grown CNTs over $\text{Al}_2\text{O}_3\text{-Fe}$ catalyst (10000 \times).

Through the SEM analysis, the $\text{Al}_2\text{O}_3\text{-Co/Fe}$ catalyst found to show excellent activity towards the growth of aligned CNTs. As depicted in Figure 3, the arrays of CNTs grown orderly with length up to 10 μm . In addition, the as-grown nanotubes possess relatively small diameter of 36 nm with narrow diameter distribution.

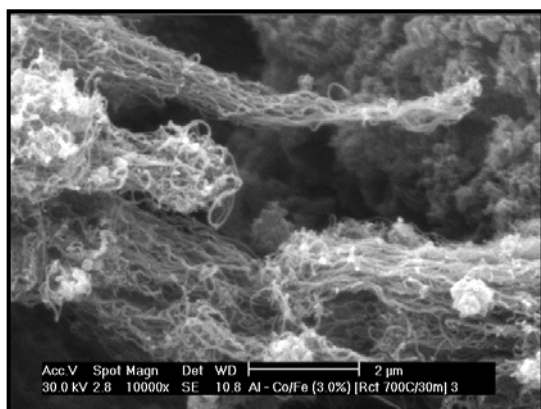


Figure 3. Aligned CNTs grown over $\text{Al}_2\text{O}_3\text{-Co/Fe}$ catalyst (10000 \times).

In Figure 4 (a), we can see that most of the as-grown CNTs over this catalyst are in alignment. The texture of the nanotubes is fluffy. It can be deduced that the purity and graphitization of the CNTs are good because there is no amorphous carbon and unwanted carbonaceous materials deposited on the sample.

From Figure 4 (b), the nanotubes were found to grow in the same direction and arranged in the form of arrays. The density of the CNTs is extremely high, the surface of the catalyst were highly saturated by the CNTs arrays. The quasi-alignment of the CNTs is clearly demonstrated in Figure 4 (b). The diameter of the nanotubes is down to 25 nm. Bunch of nanotubes grew from a cluster of catalyst particles, as pointed by a white circles. The purity and quality of the as-grown CNTs is undoubtedly superior compared to CNTs yields on $\text{Al}_2\text{O}_3\text{-Co}$ or $\text{Al}_2\text{O}_3\text{-Fe}$ catalysts. As a conclusion, all of the supported bimetallic catalysts are suitable to grow CNTs with good quality and purity.

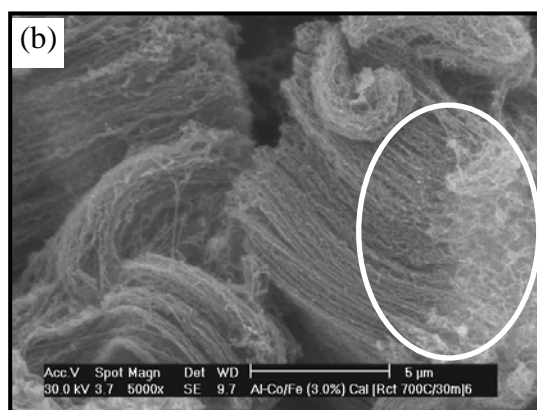
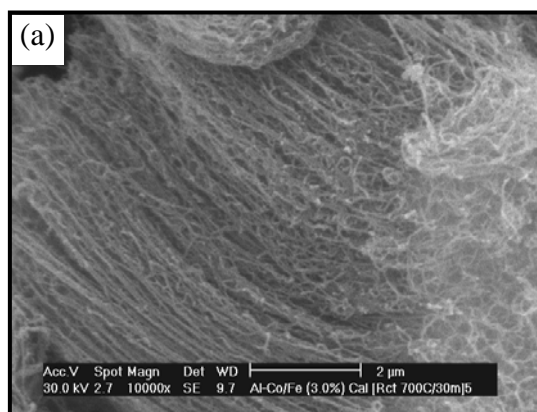


Figure 4. (a) Arrays of CNTs grown over $\text{Al}_2\text{O}_3\text{-Co/Fe}$ catalyst (10000 \times) and (b) White circle area showed bunch of CNTs grown from a cluster of catalyst particles (5000 \times).

Willems *et al.* [6] have accounted that mixture of metals greatly improve the quantity of carbon deposition with the 1:1 ratio of metals mixture giving the best yield. Yet again, the high activity of mixed metals in producing nanotubes has been proved by Willems *et al.* [7]. Moreover, the presence of aligned CNTs in the Co/Fe catalysts is in agreement with Mukhopadhyay and Mathur [8] result, which revealed that the present of both Co and Fe metal catalysts created the alignment of the CNTs. In addition, Kukovecz *et al.* [9] described that the supported Fe catalysts are efficient in producing CNTs of somewhat lower quality, as the Co catalysts produced good quality nanotubes. However, the bimetallic catalyst Co/Fe gave the best quality and quantity of CNTs.

3.2 Field-Emission Scanning Electron Microscope (FE-SEM) Analysis of As-grown CNTs

The FE-SEM micrographs of CNTs grown over $\text{Al}_2\text{O}_3\text{-Co/Fe}$ catalyst are shown in Figure 5 (a) and (b). At low resolution (Figure 5 (a)), the as-grown CNTs are observed in bundles with uniform diameter. It is clearly shown that the surface of catalyst particles was fully covered by CNT bundles, this proved that the $\text{Al}_2\text{O}_3\text{-Co/Fe}$ catalyst gave high yield of CNTs. The alignment of CNTs was not observed here as it was destroyed during the sample preparation for FE-SEM analysis.

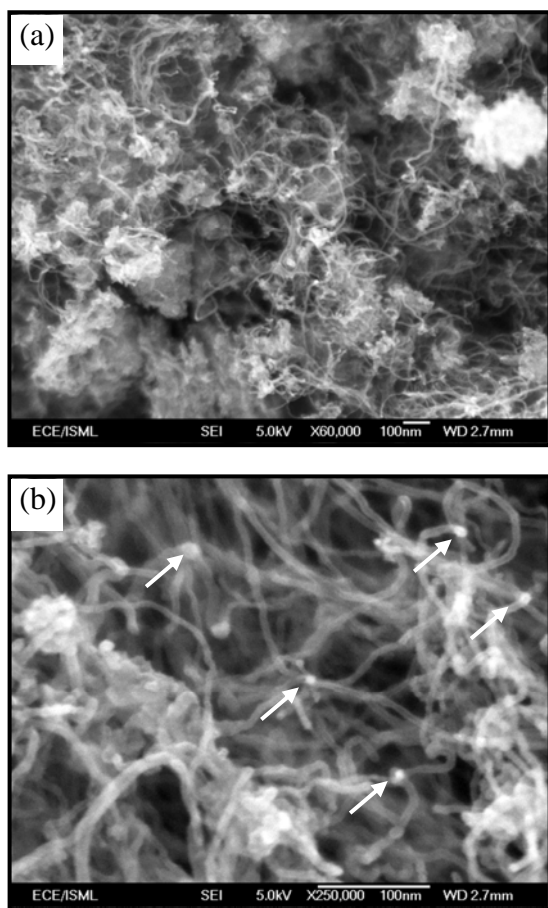


Figure 5. (a) FE-SEM image of CNTs grown over $\text{Al}_2\text{O}_3\text{-Co/Fe}$ catalyst ($60000\times$) and (b) The hollow structure of CNTs clearly observed from the micrograph of $250000\times$ magnification.

High resolution FE-SEM micrographs (Figure 5 (b)) had proven that the $\text{Al}_2\text{O}_3\text{-Co/Fe(3.0)Cal}$ catalyst is able to grow CNTs instead of carbon nanofibres. The smooth surface and the hollow structure of CNTs can be observed in the micrographs. The outer diameter of the CNT is estimated as 7.14 nm and length up to 400 nm. The growth direction of the CNTs is random and the nanotubes are entangled and twisted. It is observed that each nanotube ended by showing a white spot (as pointed by the white arrows). These white spots originated from the metal catalyst particles, which indicates that the CNTs were grown by tip growth mechanism. This observation is similar to the work of Shajahan *et al.* [10]. The substrate-catalyst interaction is weak, hence the catalyst particles are lifted up by the growing CNT and continue to promote CNT growth at its tip (tip growth mechanism), as described by Ando *et al.* [4].

3.3 Transmission Electron Microscope (TEM) Analysis of As-grown MWNTs

The TEM micrographs of as-grown CNTs obtained from $\text{Al}_2\text{O}_3\text{-Co}$ catalyst are presented in Figure 6 (a) and (b). As depicted in the low resolution image (Figure 6 (a)), the CNTs grown in random direction from the Co catalyst base

and extended with an open end. The CNTs appeared as entangled bundles. The CNTs possessed uniform diameter. The CNTs produced by $\text{Al}_2\text{O}_3\text{-Co}$ catalyst are MWNTs as observed from Figure 6 (b). The nanotube walls are turbostratic but not straight, two sides of the MWNT has different layers of wall. The inner diameter was estimated as 12.28 nm whereas the outer diameter was 30.64 nm. Some amorphous carbon was attached to the outer layer of the MWNT. However, the MWNT shows high degree of graphitization with low defect observed on the wall structure.

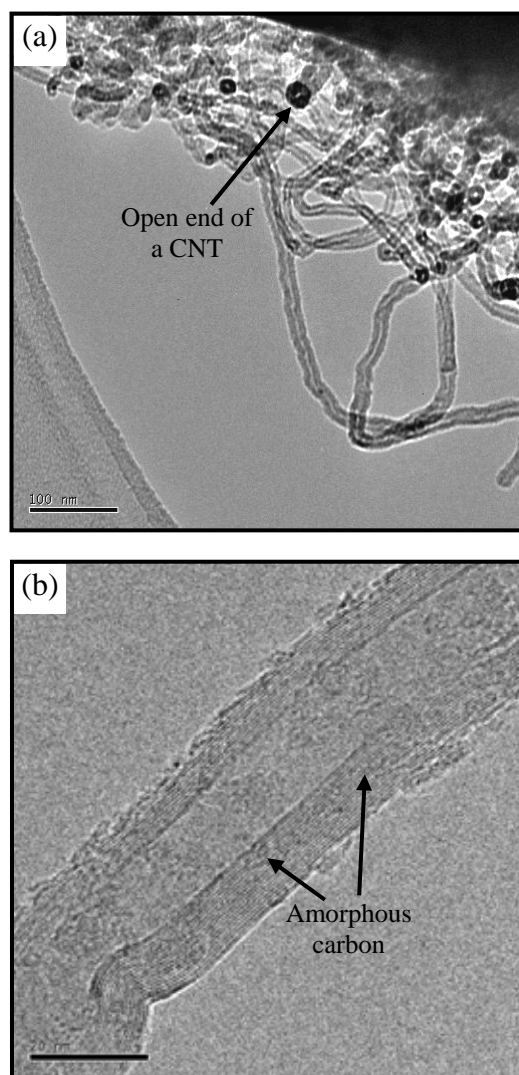


Figure 6. (a) TEM image of open ended MWNTs (scale bar: 100 nm) and (b) the turbostratic walls of MWNTs grown over $\text{Al}_2\text{O}_3\text{-Co}$ catalyst (scale bar: 20 nm).

Figure 7 represent the TEM micrographs of CNTs grown over $\text{Al}_2\text{O}_3\text{-Fe}$ catalyst. As observed from Figure 5.15 (a), this CNTs product can be classified as MWNT, although the graphite layers are not parallel. The MWNT was curved and covered by amorphous carbon at the outer layer. The MWNTs are twisted and presented as aggregates, only a few nanotubes appeared as long and straight tubes. The presence of aggregates is due to a poor graphitization process, as concluded by Willems *et al.* [7]. This MWNT

consist of 7 layers of the graphene walls with inner and outer diameter of 8.39 and 14.84 nm respectively.

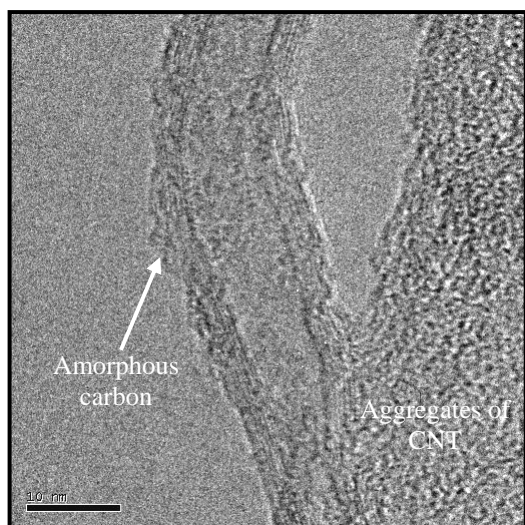


Figure 7. TEM micrograph of MWNTs grown over Al_2O_3 -Fe catalyst (scale bar: 10 nm).

The capability of Al_2O_3 -Co/Fe(3.0) catalyst to grow good quality CNTs is superior. As illustrated in Figure 8 (a), the CNTs were very long and well-graphitized with uniform diameter. There is a catalyst particle trapped at the tip of the MWNT, as pointed in the micrograph. In Figure 8 (b), it is observed a MWNT with very thin walls, it only consists of 6 layers of graphene walls. The MWNT possessed inner diameter of 4.21 nm and outer diameter of 8.16 nm. Figure 8 (c) attested that the as-grown MWNT has superb characteristic. The purity and the graphitization of the MWNT are very high. The carbon building units arranged in well order to produce extremely turbostratic and defect-free graphene layers. The inner diameter is 3.57 nm whereas the outer diameter is 11.43 nm. The distance between layers is measured as 0.33 nm, which is in agreement with inter-layer distance of graphite. Thus, the TEM analysis is again assured the high performance of the bimetallic catalysts in growing high quality MWNTs.

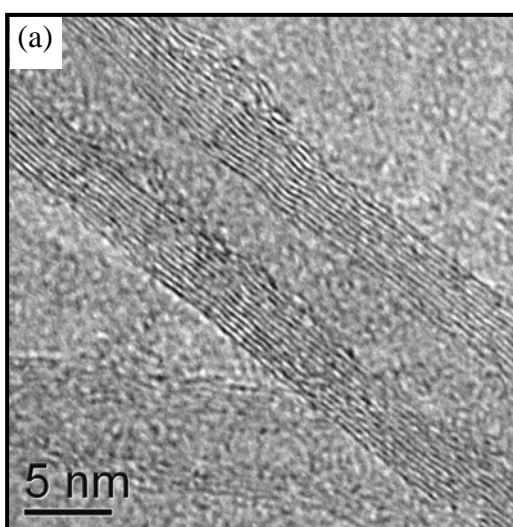
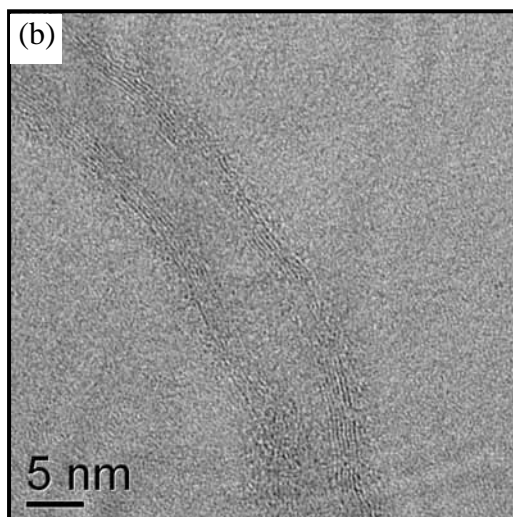
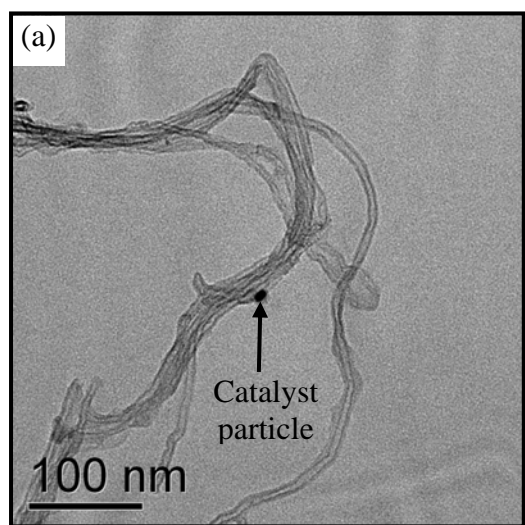


Figure 8. (a) Catalyst particle entrapped at the tip of the MWNT grown over Al_2O_3 -Co/Fe catalyst (scale bar: 100 nm), (b) MWNT with very thin walls (scale bar: 5 nm) and (c) As-grown MWNT with extremely turbostratic and defect-free graphene layers (scale bar: 5 nm).

For the synthesis of CNTs using bimetallic catalysts, the as-grown CNTs possess higher quality and degree of graphitization, as revealed in the SEM, FE-SEM and TEM analyses. As explained previously, the Co catalyst help to increase graphitization of nanotubes, whereas Fe catalyst is effective in decompose C_2H_2 for nanotubes growth. This is in agreement with Colomer *et al.* work [11], which explained that the catalysts prepared from a mixture of metals are more efficient for CNT production than those prepared with a single metal.

4. Conclusion

The Scanning Electron Microscopy (SEM), Field Emission-Scanning Electron Microscopy (FE-SEM) and Transmission Electron Microscopy (TEM) analyses gave comprehensible images of the as-grown CNTs over series of catalysts. The physical appearance of the catalysts and CNTs produced provided brief information about the

samples. All of the Al₂O₃ supported catalysts are able to grow CNTs in different configuration and density. The ability to produce CNTs was vary in the order of Co/Fe > Co > Fe catalysts. The Co catalysts produced good quality nanotubes, while the Fe catalysts are efficient of C₂H₂ decomposition in giving higher carbon deposition yield, thus, the bimetallic catalysts showed the best performance in producing high quality and quantity CNTs. The finest MWNT consists of 11 layers of turbostratic graphene wall with inner diameter of 3.57 nm and outer diameter of 11.43 nm as well as distance between layers of 0.33 nm.

Acknowledgement

The authors would like to thank MOSTI for funding this project under the IRPA mechanism of the national fuel cell project.

References

- [1] Iijima, S. 1991. Helical Microtubules of Graphitic Carbon. *Nature*. 354(6348): 56 – 58.
- [2] Dresselhaus, M.S. and Avouris, P. 2001. Introduction to Carbon Materials Research. In: Dresselhaus, M.S., Dresselhaus, G. and Avouris, P. *Carbon Nanotubes: Synthesis, Structure, Properties and Applications*. Berlin, Germany: Springer-Verlag. 1 – 9.
- [3] Kong, J. 2002. *Carbon Nanotubes: Synthesis, Integration and Properties*. Stanford University: Ph.D Thesis.
- [4] Ando, Y., Zhao, X., Sugai, T. and Kumar, M. 2004. Growing Carbon Nanotubes. *Materials Today*. Review Feature: 22 – 29.
- [5] Buang, N.A., Aziz, M., Sanip, S., Tee, J.C., Abidin, Z. and Ismail, A.F. 2005. Effect of Pretreatment of Synthetic and Natural Carbons as Starting Materials for Carbon Nanotubes. *Journal of Metastable and Nanocrystalline Materials*. 23: 285 – 288.
- [6] Willems, I., Konya, Z., Colomer, J.F., Van Tendeloo, G., Nagaraju, N., Fonseca, A. and Nagy, J.B. 2000. Control of the Outer Diameter of Thin Carbon Nanotubes Synthesized by Catalytic Decomposition of Hydrocarbons. *Chemical Physics Letters*. 317: 71 – 76.
- [7] Willems, I., Konya, Z., Fonseca, A. and Nagy J.B. 2002. Heterogeneous Catalytic Production and Mechanical Resistance of Nanotubes Prepared on Magnesium Oxide supported Co-based Catalysts. *Applied Catalysis A: General*. 229: 229 – 233.
- [8] Mukhopadhyay, K. and Mathur, G.N. (2003). Synthesis of 2D Quasi-Aligned Multiwalled Carbon Nanotubes by Catalytic Chemical Vapor Deposition Method. *Journal of Nanoscience*. 2(3): 153 – 164.
- [9] Kukovecz, A., Konya, Z., Nagaraju, N., Willems, I., Tamasi, A., Fonseca, A., Nagy, J.B. and Kiricsi, I. (2000). Catalytic Synthesis of Carbon Nanotubes over Co, Fe and Ni Containing Conventional and Sol-Gel Silica-Aluminas. *Physical Chemistry Chemical Physics*. 2: 3071 – 3076.
- [10] Shajahan, Md., Mo, Y.H., Kibria, A.K.M.F., Kim, M.J. and Nahm, K.S. (2004). High Growth of SWNTs and MWNTs from C₂H₂ Decomposition over Co-Mo / MgO Catalysts. *Carbon*. 42: 2245 – 2253.
- [11] Colomer, J.F., Bister, G., Willems, I., Konya, Z., Fonseca, A., Van Tendeloo, G. and Nagy, J.B. (1999). Synthesis of Single-Wall Carbon Nanotubes by Catalytic Decomposition of Hydrocarbons. *Chemical Communications*. 14: 1343 – 1344.

Residential Wastewater as a Source of Competent Phenol Degrading Consortia

Norhani Jusoh and Firdausi Razali*

*Department of Bioprocess Engineering, Faculty of Chemical & Natural Resources Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5535805, Fax: +60-7-5581463, E-mail: noniey82@yahoo.com*

Abstract

Excellent phenol-degrading microbes are mainly characterized by their robustness to adapt at high phenol concentrations. The study aimed to evaluate the tolerance of acclimatized phenol degrading consortia from residential wastewater in continuous cultures. Initially, the mixed culture was enriched to increase the microbe's survival rate, so that significant reduction of phenol could be reliably quantified. Batch-mode acclimatization was then executed by exposing the culture with the highest possible phenol tolerance level, namely 1200 mg/L, for two weeks to test its endurance and adaptability. Surprisingly, the culture was demonstrated 93% of phenol removal with 3.2 mg/L/hr of phenol degradation rate. Continuous degradation in 1.5 L working volume of chemostat was performed by loading a 1000 mg/L phenol at incremental rate of 0.5, 1.0, 1.4, 1.8, and 2.2 ml/min. The increment of loading rate was justified once constant biomass optical density was observed. The acclimatized culture, maintained the average phenol level at 316 mg/L in the continuous mode throughout 121 hours of biodegradation. Results obtained showed that acclimatization has insignificant effect on continuous phenol biodegradation. However, outcomes from this study provide strong evidence that residential wastewater treatment could be a potential source of phenol degrading consortia to solve phenol contamination problem in the environment, which has been closely associated with pesticide, textile, and oil refinery industries.

Keywords:

Residential Wastewater, Phenol, Acclimatization

1. Introduction

Phenol has been declared as hazardous pollutant [1] and this compound is highly toxic even at low concentration [2]. Phenol is normally found in wastewaters from coal carbonization plant, oil refineries, petrochemical industries, phenol based polymerization processes, pharmaceutical, paints and varnish producing units [3]. Existing phenol removal methods are including adsorption on to different matrices, solvent extraction, chemical oxidation and irradiation [4]. Biological treatment has demonstrated to be effective in phenolic wastewater treatment. This method is preferable because economical [5] and production of innocuous end products [1]. Bioremediation is an example of biological method that had been widely studied in treating phenolic compounds. Bioremediation uses naturally occurring microorganisms to degrade hazardous substances [6].

Many researches on phenol bioremediation have been done involving many species of microorganisms. Microorganisms that normally used in phenol biodegradation are including *Pseudomonas putida* [2, 3, 7, 8], *Candida tropicalis* [1] and *Acinetobacter* sp. [9]. This study has used residential wastewater as source of microorganism to degrade phenol. Residential wastewater

is choosing because containing wider spectrum of microorganisms as well as metabolic properties. Biodegradation of organic chemicals by microorganisms using pure cultures usually can produce toxic intermediates. Therefore, mixed culture is preferable in this study because can overcome this problem [7].

Phenol biodegradation by mixed culture has been carried out by continuous biodegradation. Continuous biodegradation is preferable because this mode of cultivation is mimicked to the real industrial scenario in order to better understand the basic mechanisms of microbial removal in wastewater treatment plant. Hence, continuous biodegradation of phenol has been studied by performing stepwise increased in dilution rates [8, 10]. Stepwise increased in dilution rates was performed in order to examine the unsteady state behaviour of acclimatized cultures that during continuous cultures, were subjected to abrupt changes in their operating conditions, which is increasing dilution rate. Straube and Hensel [10] observed that the stability of the system decreased with increasing dilution rate and increasing phenol inlet concentration.

Acclimatization was proved to overcome substrate inhibition problems that usually occurred in biodegradation at high concentration of phenol [11]. During acclimatization stage, certain enzymes in the microorganism are induced, hence this microbes are available for next metabolism reaction [12]. The objective of this study to develop a quick, reliable, reproducible, and robust bioremediation technique that capable of degrading common industrial

* Corresponding Author. E-mail: firdaus@fkkksa.utm.my,
Tel : +60-7-5535513, Fax: +60-7-5581463

toxic liquid waste. Monod equation [13] has been used to describe kinetic data of phenol degradation by residential microorganisms.

2. Materials and Methods

2.1 Experimental Setup

Batch and continuous biodegradation was performed in a bioreactor (B.BRAUN, Germany) with a working volume of 2L equipped with pH, agitation, temperature and dissolved oxygen probe. The agitation rate was 400 rpm to maintain a homogeneous culture. Dissolved oxygen concentration was controlled in the range of 2-4 L/min by air sparging using a BEBICON compressor. For continuous system, an external multi channel pump (Cole Parmer, USA) was used to feed and pump out the effluent from the bioreactor.

2.2 Culture and Media

Residential wastewater from Desa Bakti Wastewater Treatment Plant had been used as source of microorganisms. The wastewater sample was filtered to remove visible suspended solid and stored at 4°C. Microbial consortium use in this study was developed as describe latter. The Luria-Bertani (LB) medium used for the enrichment stage consist of 5g/L yeast extract (CONDA, Spain), 10 g/L Tryptone (CONDA, Spain), 15 g/L glucose (Sigma Chemical Co., USA) and 5 g/L sodium chloride (NaCl) (MERCK, Germany). Phenol (Scharlau Chemie S.A, Spain) was used in this research as toxic compound.

2.3 Culture Enrichment

The experimental procedure was initially run in batch mode. Desa Bakti wastewater was added into 2 L bioreactor with 1.5 L of working volume. The enrichment stage was initially carried out for ten hours. LB medium consist of 7.5 g of yeast extract, 15 g of tryptone, 7.5 g of NaCl and 22.5 g of glucose was added at the beginning of the experiment. The growth of microorganism was determined by optical density (OD) reading at 600 nm. The enrichment stage was performed until stationary growth phase was achieved.

2.4 Acclimatization

The enriched culture was initially acclimatized to the phenol before biodegradation stage is carrying out. During this stage, bioreactor was operated in batch mode. The bioreactor that contained residential microorganisms was filled up with phenol at concentration level of 1000mg/L. No nutrients were added during this process and air was supplied to ensure the dissolved oxygen was enough for this stage.

2.5 Continuous Biodegradation

Enriched microorganisms were used to inoculate the bioreactor. Continuous pumping of phenol solution was started after the culture had entered the stationary growth phase for microorganisms that was not adapted to phenol. For the microorganisms that were acclimatized to phenol,

continuous feeding will be started after complete degradation of phenol in acclimatization stage. No nutrients were added and air was supplied continuously [2].

2.6 Analytical Methods

Periodic samples of the mixed liquor were taken throughout all the experiments in order to measure the concentrations of microorganism. The OD was measured spectrometrically at 600nm in 1 cm³ cuvettes. As a rule of thumb, an optical density of 1 unit is corresponds to approximately 1 g/L of biomass cell.

Calorimetric method was used to determine phenol concentration by using Folin –Ciocalteu reagent. 2 mL sample was added with 0.3 mL of sodium carbonate (200g/L) and 0.1 mL of Folin-Ciocalteu reagent. After 1 hour of incubation the absorbance were measured at 750 nm [14]. A standard curve was prepared with concentration rages from 0.002 to 0.01 g/L.

A series of mass balance calculation has been carried out for all continuous biodegradation in the purpose of to determine the phenol concentration when biodegradation did not occurred. This calculation will be defined as calculated phenol concentration throughout this work.

3. Results

3.1 Continuous Biodegradation with Acclimatization

Residential wastewater firstly enriched in bioreactor to increased microorganisms concentration. All the growth curves obtained show lag phase occurred immediately after inoculation started. After 6 hours of enrichment stage, cell started to multiply exponentially with times. After about 24 hours of enrichment, OD was increased to 4.677. Growth of microorganisms then decelerated due to either depletion of essential nutrients and the microorganisms had entered stationary phase where the cell division is zero and the growth rate is equal to death rate. The enriched microorganism was then acclimatized to the 1000mg/L of phenol. Figure 1 represents the bioreactor condition during acclimatization of residential microorganisms to phenol compounds.

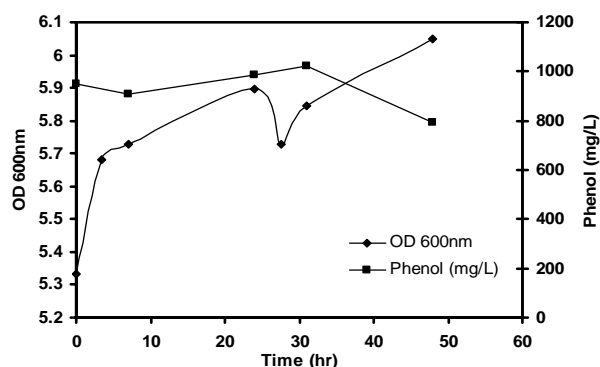


Figure 1: Phenol degradation during two days of acclimatization

During 48 hours of batch acclimatization, the concentration of microorganisms merely constant with specific growth rate, $\mu=0.003 \text{ h}^{-1}$. This value of μ is too small compared to 0.174 h^{-1} during enrichment. Low oxygen concentration (1.2 to 8.1%) shows that aerobic biodegradation occurred and resulted in phenol reduction. Within 48 hours of acclimatization, 21% of phenol had been degraded. The curve shows that the microorganism still can grow although no nutrient was supplied during this stage. The microorganisms had degraded phenol and used this substance as source of carbon.

This acclimatized microorganism was then used for continuous biodegradation of phenol and the results are shown in Figure 2.

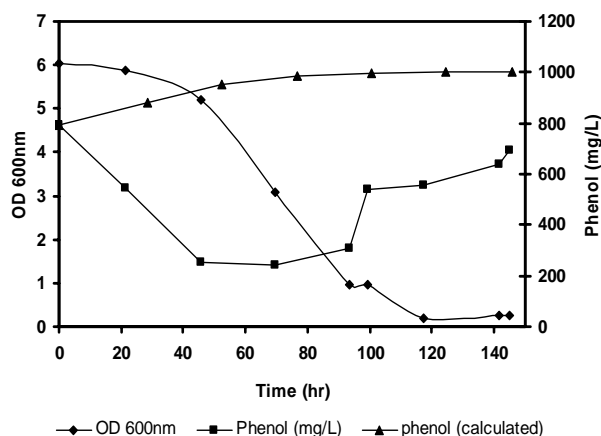


Figure 2: Continuous biodegradation after two days of acclimatization

Continuous biodegradation was started with initial feeding of 0.5 ml/min of phenol to the acclimatized microorganisms. After 24 hours, when the cell concentration was nearly constant, the feed flow rate of phenol, F was increased to 1.0 ml/min. This process was repeated by using other feed flow rates 1.4, 1.8 and 2.2ml/min. The microorganisms degraded phenol to the lowest concentration of 242.66 mg/L but then the phenol concentration increased due to the increased in the dilution rate. The average phenol concentration level during the continuous treatment is 508 mg/L. These microbes are believed to reduce 48 % from total phenol when compared to the system in the absent of microorganism by mass balance calculation. Biodegradation occurred but decreased with time due to reduction of microorganism concentration.

Continuous biodegradation after two days of acclimatization results in maximum growth rate, $\mu_{\max} = 0.053 \text{ h}^{-1}$ and substrate utilization constant, $K_s=44.7 \text{ mg/L}$. This growth rate value is lower, compared to 0.34 h^{-1} found by Lallai [15], studied about aerobic mixed culture. K_s value, 44.7 mg/L is higher than stated in literature, 1.61 mg/L [15].

Figure 3 shows the dependence of OD and phenol concentration on dilution rate during continuous biodegradation with critical dilution rate, $D_c=0.042 \text{ h}^{-1}$. Microorganisms degraded the phenol when the dilution

rate, D is low ($0.02-0.056 \text{ h}^{-1}$). As the dilution rate was increased to 0.072 h^{-1} , the microorganisms cannot degrade the phenol anymore and resulted in increase of phenol concentration. However, the phenol concentration is still below the phenol calculated level.

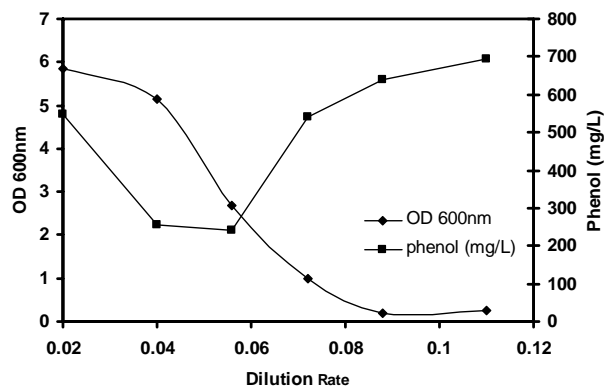


Figure 3: The dependence of OD and phenol concentration on dilution rate

Experiment then was carried out to study effect of longer acclimatization period on continuous biodegradation of phenol. Figure 4 shows the degradation of phenol during 14 days of acclimatization.

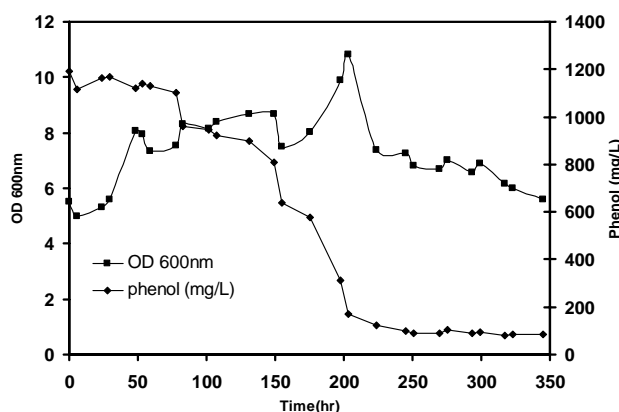


Figure 4: Phenol degradation during 14 days of acclimatization

During 14 days of batch cultivation, 93% of phenol reduction proved that microorganisms had successfully degraded phenol compound with 3.2 mg/L/hr of phenol degradation rate. As shown in Figure 4, microorganisms concentration increased with the reduction of phenol concentration in the beginning of the experiment. However, after around 8.5 days of degradation, microorganisms concentration decreased as more than 80% of phenol substances was degraded. Decreasing in microorganisms concentration might be due to the low concentration of carbon source, namely phenol. This degradation results in growth rate of 0.007 h^{-1} and K_s value of 0.15 mg/L.

The acclimatized microorganism with 70 mg/L of initial phenol concentration then was continuously supplied

with 0.5 ml/min of 1000mg/L of phenol. After 120 hours of continuous biodegradation, results in average phenol concentration at 316 mg/L with 61 % of phenol reduction and the results were illustrated in Figure 5.

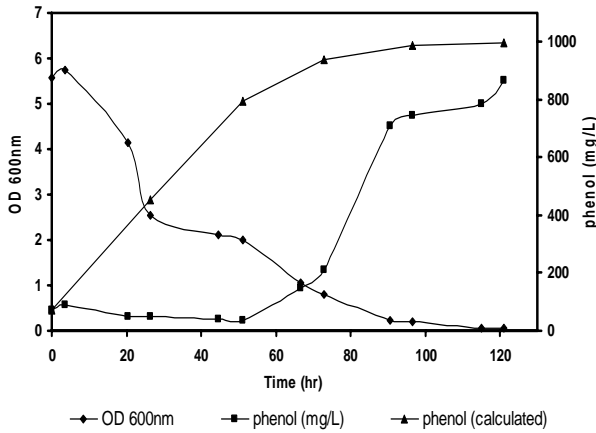


Figure 5: Continuous biodegradation after 14 days of acclimatization

The concentration of microorganism was slowly decreased with $\mu_{max} = 0.0822 \text{ h}^{-1}$ and $K_s=173 \text{ mg/L}$. The system is initially stable and phenol degradation occurred when F is 0.5 and 1mL/min. When F was 1.4 mL/min, the phenol degradation slower and results in phenol increased to 210 mg/L. When F was 1.8 mL/min, phenol concentration increased to 744 mg/L. However, during the experiment the microorganisms successfully maintained phenol concentration below the calculated phenol level.

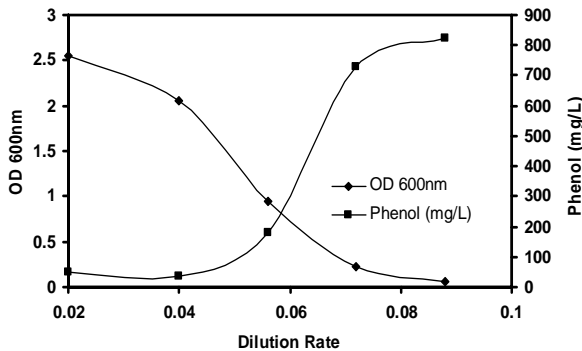


Figure 6: The dependence of OD and phenol concentration on dilution rate

Figure 6 shows that, 14 days of acclimatization results in 0.04 h^{-1} of D_c . This value is lower compare to 0.077 found by Hensel and Straube [10] for continuous phenol biodegradation by *Rhodococcus sp.* P1.

3.2 Continuous Biodegradation without Acclimatization

For this study, enriched microorganism with initial OD of 5.296 was continuously supplied with phenol at concentration of 1000 mg/L as shown in Figure 7.

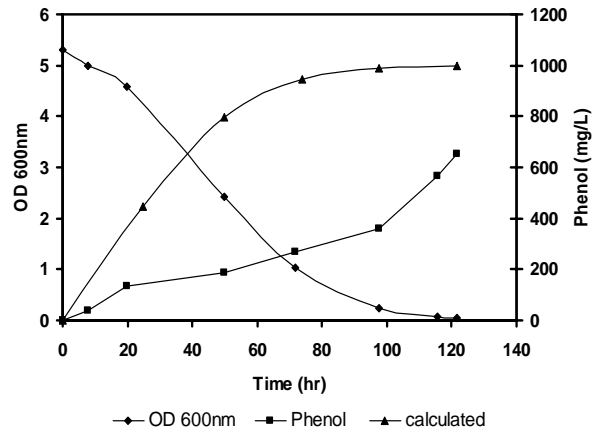


Figure 7: Continuous biodegradation without acclimatization

The microorganisms were firstly fed with 0.6 ml/min of phenol. When the cell concentration was nearly constant, the feed flow rate of phenol was increased to 1.0 ml/min as previous experiments. The non acclimatized microorganisms have degraded the phenol and results in 67 % of phenol reduction. The graph shows that, the mixed cultured still maintained the phenol level below the calculated phenol level when biodegradation did not occurred.

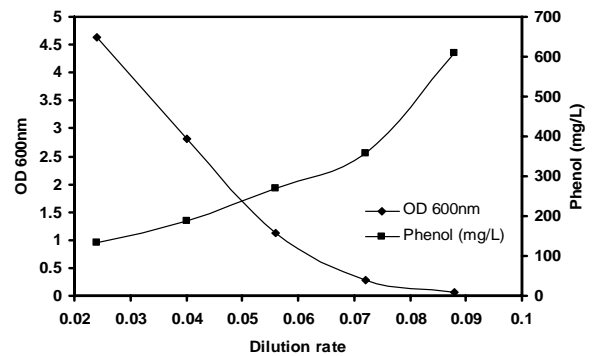


Figure 8: The dependence of OD and phenol concentration on dilution rate

Figure 8 shows that as the dilution rate increased, phenol concentration also increased to 650.83 mg/L with average phenol concentration of 276 mg/L.

3.3 Effect of Acclimatization on Continuous Biodegradation

In many cases, acclimatization of microorganisms to the toxic substances before full treatment seems to be an effective approach and enhance the biodegradation process. However, in this study all system have merely same amount of reduction which is between 48 to 67% of phenol reduction. In term of phenol reduction, the difference seems insignificant between acclimatized and non acclimatized microorganisms.

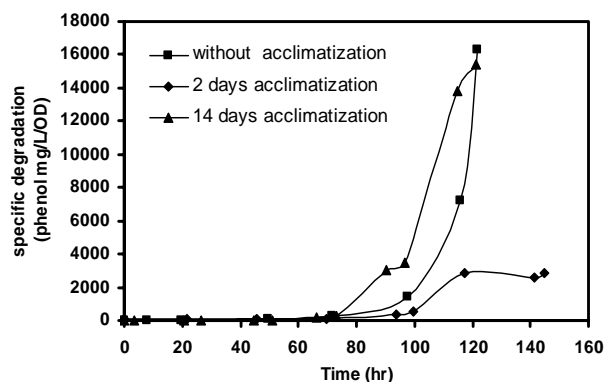


Figure 9: Effect of acclimatization on specific degradation

Figure 9 shows effect of acclimatization on specific degradation of phenol in continuous system with stepwise increase in dilution rates. Nevertheless, in term of specific degradation, continuous biodegradation without acclimatization stage and 14 days of acclimatization has merely the same value of specific degradation.

4. Discussion

In batch biodegradation of 1200 mg/L of phenol, complete biodegradation is nearly achieved with small increase in the growth rate. This result is comparable with result found by Gonzalez *et al.*, [2] with complete reduction of phenol after 10 days of degradation with 1000 mg/L of initial concentration. This result proved that, mixed culture from residential wastewater showed the similar performance as pure culture.

Mixed culture from residential wastewater was able to degrade phenol in the continuous system in the rates up to 67%. With increasing dilution rate the stability of the system decreased. When the dilution rate is increased, phenol slowly accumulated in the medium resulting in toxic effects and washout of the culture. This condition was a typical result of substrate inhibition by the toxic substrate. Results obtained are agreed with Sokol and Migiro [8], stated that disturbance may cause a transient increase in substrate concentration, washout of the biomass and also the biomass concentration in the bioreactor will continuously decrease to zero and consequently the substrate concentration in the bioreactor will increase until it reaches substrate concentration in the feed.

However, for this study, acclimatization did not give significant effect to continuous phenol biodegradation. The non acclimatized microorganism has higher μ_{max} value 0.68 h^{-1} , compared to acclimatize microorganism, 0.053 h^{-1} . K_s value, 3024 mg/L for non acclimatized microorganisms is too high compared to acclimatized microorganism 44.7 mg/L. This means that, acclimatized microorganisms have higher affinities for the substrate than non acclimatized microorganisms. D_c value for non acclimatized microorganisms is less than 0.024 h^{-1} , lower compared to acclimatized microorganism, 0.042 h^{-1} . Therefore,

acclimatization can be excluded for phenol biodegradation by residential wastewater.

5. Conclusions

The stability of the continuous system decreased with increasing dilution rate. Initially, microorganisms can withstand to the disturbance. However when dilution rate is too high, phenol accumulated in the medium and resulting in toxic effects and lead to culture washout. All system has merely same value of phenol reduction, which is around 50%. Although the difference is insignificant between acclimatized and non acclimatized microorganisms performances, these microorganisms still successfully maintained the phenol concentration below the calculated phenol level. The experiment results showed that it is possible to treat wastewater containing phenol in continuous biodegradation.

Acknowledgements

The authors are thankful to Mr. Syed Annur Fauad, lecturers, technicians and students in Department of Bioprocess Engineering for their contribution in fulfilled this study. This study was funded by UTM Short Term Grant, vote no: 75200 and UTM-PTP Scholarship.

References

- [1] Chen, K.C., Lin, Y.H., Chen, W.H. and Liu, Y.C. 2002. Degradation of Phenol by PAA-immobilized *Candida tropicalis*. *Enzyme and Microbial Technology*. 31:490-497.
- [2] Gonzalez, G., Herrera, M.G., Garcia, M.T. and Pena, M.M. 2000. Biodegradation of Phenol in Continuous Process: Comparative Study of Stirred Tank and Fluidized Bed Reactors. *Bioresource Technology*. 76: 245-251.
- [3] Bandhyopadhyay, K., Das, D. and Bhattacharyya, P. and Maiti, B.R. 2001. Reaction Engineering Studies on Biodegradation of Phenol by *Pseudomonas putida* MTCC 1194 Immobilized on Calcium Alginate. *Biochemical Engineering Journal*. 8: 179-186.
- [4] Spiker, K., Crawford, L., and Thiel, C., 1992. Oxidation of Phenolic and Non Phenolic Substrates by the Lignin peroxidase of *Streptomyces viridosporous* T7A. *Applied Microbial Biotechnology*. 37: 518-523.
- [5] Kim, J.H., Oh, K.K., Lee S.T, Kim S.W. and Hong S.I. 2002. Biodegradation of Phenol and Chlorophenols with Defined Mixed Culture in Shake Flasks and a Packed Bed Reactor. *Process Biochemistry*. 37: 1367-1373.

- [6] U.S. Environmental Protection Agency. A Citizen's Guide to Bioremediation. April 1996. Washington DC. EPA 542-F-96-007.
- [7] Monteiro, A.A.M.G., Boaventura, R.A.R. and Rodrigues, A.E. 2000. Phenol Biodegradation by *Pseudomonas putida* DSM 548 in a Batch Reactor. *Biochemical Engineering Journal*. 6: 45-49.
- [8] Sokol, W. and Migiros, C.L.C. 1995. Controlling a Continuous Stirred Tank Bioreactor Degrading Phenol in Stability Range. *The Biochemical Engineering Journal*. 62: 67-72.
- [9] Hao, O.J., Kim, M.H., Seager, E.A. and Kim, H. 2002. Kinetics of Phenol and Chlorophenol Utilization by *Acinetobacter* species. *Chemosphere*. 46: 797-807.
- [10] Hensel, J. and Straube, G. 1990. Kinetic Studies of Phenol Degradation by *Rhodococcus* sp. P1 II. Continuous Cultivation. *Antonie van Leeuwenhoek* 57: 33-36.
- [11] Lob, K.C. and Tar, P.P. 2000. Effect of Additional Carbon Source on Biodegradation of Phenol. *Bulletin of Environmental Contamination and Toxicology*. 64: 756-767.
- [12] Kumar, A., Kumar, S. and Kumar S. 1994. Biodegradation Kinetics of Phenol and Catechol Using *Pseudomonas putida* MTCC 1194. *Biochemical Engineering Journal*. 22 (2005) 151-159.
- [13] Stanbury, P.F. and Whitaker, A. 1984. *Principles of Fermentation Technology*. 1st edition. Pergamon Press Ltd. United Kingdom.
- [14] Box, J.D. 1981. Investigation of the Folin –Ciocalteu Phenol Reagent for the Determination of Polyphenolic Substances in Natural Waters. *Water Research*. 17:511-524
- [15] Lallai, A. and Mura, G. 1994. Steady-state Behaviour of an Aerobic Mixed Culture Growing on Phenol in a Continuous Stirred Reactor. *The Chemical Engineering Journal*. 53:B47-B55

Vapor-Liquid Equilibrium of Ethanol-Water System: Effect of Ultrasonic Waves

Siti Kholijah Abdul Mudalip¹, Adnan Ripin^{1*}, Rosli Mohd Yunus²

¹Chemical Engineering Department, Faculty of Chemical & Natural Resources Engineering,
Universiti Teknologi Malaysia, 81300 Skudai, Johor Bahru, Malaysia.
Tel: +60-7-5535690, Fax: +60-7-5581463, Email: ctkhol@yahoo.com, adnan@fkkksa.utm.my

²Faculty of Chemical & Natural Resources Engineering,
Kolej Universiti iKejuruteraan and Teknologi Malaysia, Kuantan, Pahang, Malaysia.
Tel: +6-019-2401134, Email: rmy@kuktem.edu.my

Abstract

Separation of ethanol and water is one of the challenging processes in chemical industry due to azeotropic formation in the mixture. The conventional method of separating the mixture is by introducing third component (solvent) which eliminate the azeotropic formation. However, the method requires additional unit operations to recover the solvent. An innovative method of incorporating ultrasonic wave fields in distillation was proposed in this study. The effect of different ultrasonic power, 50, 100, 200 and 250 Watts, at frequency of 40 kHz on vapor-liquid equilibrium (VLE) of ethanol-water system was investigated. The finding shows that the VLE of the system was favorably altered in the presence of the ultrasonic waves. The shifting of the azeotropic point was directly related to the ultrasonic power. In this study, at input power of 100 Watts, the ultrasonic wave fields manage to totally eliminate the azeotropic point. In general, the use of ultrasonic waves modifies the relative volatility and VLE of binary mixture through mechanism known as cavitation. The local shock heating which occurred during cavitation, supply an additional heating to the system, increase its turbulency and at the same time increase vaporization and mass transfer rate between phases (vapor and liquid). Due to differences in physical and chemical properties of the individual component in the mixture, the amount of ultrasonic energy being absorbed by the component in the azeotropic mixture varied accordingly. The volatile component in the azeotropic mixture, offers the most likeable spot for cavitation micro bubbles occurrence, which contain vapor of volatile component, thus make the separation possible.

Keywords:

Vapor-liquid equilibrium, ethanol-water, ultrasonic power, cavitation, relative volatility.

1. Introduction

Distillation is a method of separation based on differences in composition between liquid and vapor form from it. In chemical and petrochemical industries, distillation is one of the largest energy consumers among the several processing operations employed [1]. Distillation, in some cases, however is not the most appropriate separation techniques especially for separation of liquid mixtures with low relative volatility, or when azeotrope are form [2]. Separation of low relative volatility mixtures using normal distillation column will produce low purity product. Besides that, it also requires a prohibitive cost investment in plant and machinery [1]. An alternative technique based on addition of entrainer in mixtures have been developed and known as azeotropic distillation and extractive distillation [3, 4, 5, 6].

Distillation has been utilized as separation tools in ethanol purification for thousands of year. However, purification of ethanol from water using distillation process faced problem due to azeotropic formation at 95 wt.% of ethanol [3]. The conventional method for recovering anhydrous ethanol from fermentation broth is at least required three processing steps: (1) distillation of dilute alcohol which containing 10% of ethanol, is concentrated to its azeotropic point (95 wt.% of ethanol); (2) extractive or azeotropic distillation of the ethanol produced in the first distillation step to produced anhydrous ethanol (this step involves the addition of a third component, either an organic solvent or a strong salt, to alter the relative volatility of mixture around the azeotropic point); and (3) separation of water from the third component so that it can be recycled [7]. Literatures reported that this method consumes 50-80% of energy used in a typical fermentation ethanol manufacturing process [8].

The recent interest in searching for economical process to obtain high purity anhydrous ethanol leads researchers to explore the new dimension which can solve the azeotropic problem. Due to the importance of VLE in designing distillation column many literatures has reported the vapor-

* Corresponding Author. Email: adnan@fkkksa.utm.my, Tel: +60-7-553569, Fax: +60-7-5581463.

liquid equilibrium studies of ethanol-water system which affected by salt and others solvent or entrainers. All those proposed method rely on some additional mechanism to further modify the relative volatility of the ethanol-water system [9]. On the other hand, many processes in various technologies are improved and accelerated by the introduction of ultrasonic wave. For example, ultrasonic is used to accelerate heat and mass transfer in variety of food processing and have been successfully used to improved drying, mixing, and extraction [10, 11, 12]. Ultrasonic exposure is also a well suited method for emulsification of partly miscible liquid phases. Currently, ultrasonic is used as a homogenizer to enable determination of vapor-liquid-liquid-equilibrium (VLLE) data in systems with limited miscibility in the liquid phase [13, 14].

The objective of this study is to investigate the further effect of ultrasonic wave on VLE of the ethanol-water mixtures using vapor-liquid equilibrium apparatus which equipped with ultrasonic generating apparatus. The results obtained shows the applicability of using ultrasonic wave to alter the VLE data and finally solve the azeotropic problem occurs during separating ethanol-water mixtures

2. Methodology

2.1 Materials

Ethanol used in this study was supplied by R&M Chemical Industries (M) Sdn Bhd. The purity of ethanol used is 99.7%. Water used in this study was distilled water. The refractive indices and boiling points of each materials used were measured, and the result with those reported in the literature are listed in Table 1.

2.2 Procedure

The VLE measurements were carried out using ultrasonic vapor-liquid equilibrium apparatus developed in this study. This apparatus consist of distillation flask, condenser, water bath, ultrasonic generating equipment and thermocouples. The apparatus used was designed for a charge of 250 cm³ mixtures and can be operated at low or moderate pressure. The distillation flask was immersed in the water bath equipped with ultrasonic transducer (40 kHz) and heater. The ultrasonic transducer, supplied by Crest Ultrasonic (M) Sdn Bhd, was connected to 500 Watt ultrasonic generator. A heater, equipped with temperature controller, was used to provide the energy for boiling processes in the distillation flask. The liquid and vapor temperature was measured using thermocouples, TC-08 with precision of 0.01°C. The data were transferred using Pico data logger to a computer. The vapor and liquid phases reach equilibrium once the vapor temperature remained constant for a period of 10-20 minutes. When the equilibrium was reached, sample from distillation flask (liquid) and vapor condensed phase were taken and their refractive indices (binary mixtures) were measured using

Atogo Refractometer at 28°C with uncertainty of 1°C. This experiment was firstly done at atmospheric condition without the presence of ultrasonic wave. Then, it was repeated using different ultrasonic power of 50, 100, 200 and 250 Watts, respectively. During the experimental study, mixtures with different composition of ethanol-water (17%, 27%, 40%, 79% and 88% mole ethanol) were prepared and fed into the distillation flask.

2.3 Analysis Procedure

Atogo Refractometer was used to measure the ethanol composition in water. 10 mL mixtures of ethanol and water ranging from 0.0 to 1.0 mole fraction of ethanol were firstly prepared to generate the standard curve for the analysis of the samples. The refractive indices of each standard and pure liquid were plotted against the compositions of solutions expressed in mole fraction of ethanol. The refractive indices of each sample collected from distillate and boiler were then measured. The ethanol composition in vapor and liquid were then determined by interpolation from the plotted graph.

3. Results and Discussion

3.1 VLE study without sonication

The experimental VLE results for the ethanol (1)-water (2) system without the presence of ultrasonic wave at atmospheric pressure was firstly compared with the literature cited [6]. This was done to verify the reliability of the developed system to obtain VLE data. As can be seen in Figure 1, the VLE data obtained from this work are almost concurring with those found in the literature. This proved the practical feasibility of using the developed apparatus to obtain VLE data and also to study the effect of ultrasonic wave on VLE of binary mixtures. The experimental data, vapor and liquid compositions, boiling temperatures and calculated value of liquid activities coefficient and relative volatilities are shown in Table 2. The activity coefficients in liquid phase, γ_i were calculated using equation (1) below:

$$\gamma_i = y_i \cdot P / x_i \cdot P_i^{sat} \quad (1)$$

where the P_i^{sat} were calculated using the Antoine equation given in Table 3 along with the Antoine constant taken from the reference [16] by assuming the ideal behavior in vapor phase. The calculated γ_i shows that the ethanol (1)-water (2) systems presents a positive deviation from the ideality. Positive deviation occurs when value of $\gamma_i > 1.0$. The relative volatilities of mixture were calculated using following equation:

$$\alpha_{12} = \frac{(y_1 / x_1) / (y_2 / x_2)}{(\gamma_1 P_1^{sat}) / (\gamma_2 P_2^{sat})} \quad (2)$$

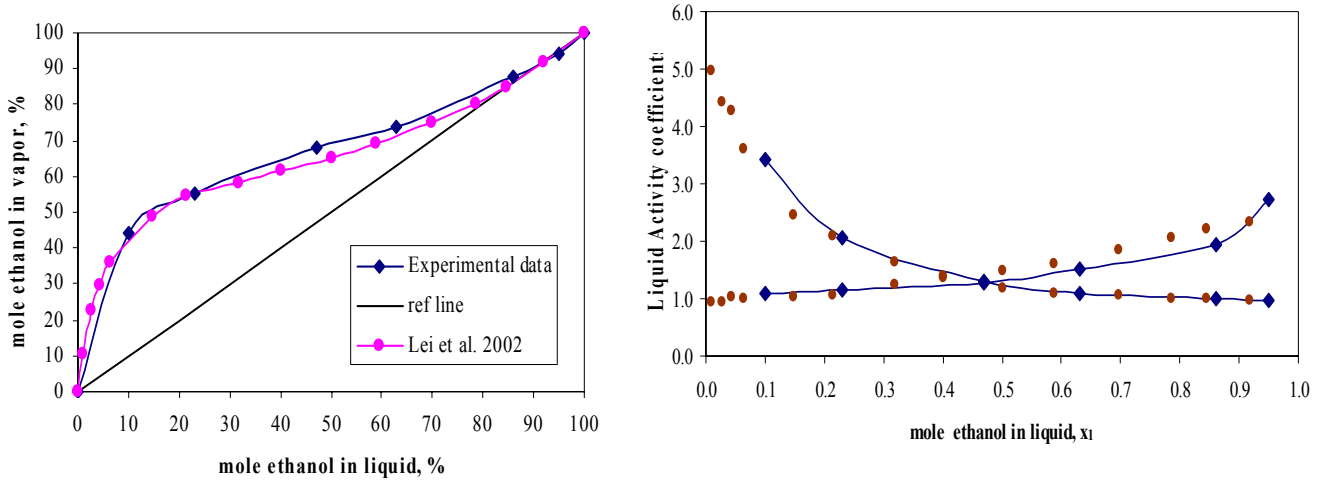


Figure 1: VLE data of ethanol (1) - water (2) at atmospheric pressure. (♦) Experimental data; (●) Literature value [5].

Table 1: Physical properties of materials

Material	Refractive indices, $n(D)$		Boiling point, T_b (K)	
	Experimental	Literature	Experimental	Literature
Ethanol	1.3580	1.3592 [15]	351.63	351.44 [15]
Water	1.3330	1.33301 [9]	373.15	373.13 [9]

Table 2: VLE data for ethanol (1) - water (2) system at atmospheric pressure

T (°C)	y ₁	x ₁	log P ₁ ^{sat}	log P ₂ ^{sat}	P ₁ ^{sat} (torr)	P ₂ ^{sat} (torr)	γ ₁	γ ₂	α ₁₂
78.4	1	1	2.8826	2.5213	763.0662	332.1420			
78.45	0.94	0.95	2.8834	2.5222	764.5758	332.8232	0.9836	2.7402	0.8246
78.82	0.88	0.86	2.8898	2.5288	775.8246	337.9007	1.0024	1.9279	1.1938
80.07	0.74	0.63	2.9111	2.5509	814.8532	355.5402	1.0955	1.5021	1.6716
80.69	0.68	0.47	2.9216	2.5618	834.8099	364.5733	1.3172	1.2586	2.3963
82.08	0.55	0.23	2.9450	2.5861	881.0340	385.5296	2.0628	1.1521	4.0918
84.78	0.44	0.1	2.9899	2.6326	976.9174	429.1415	3.4230	1.1019	7.0714
100	0	0	3.2289	2.8809	1693.8320	760.0864			
								α ₁₂ (average)	2.8749

Table 3: Antoine constant of material used in experimental study [16]

Components	A	B	C
Ethanol	8.1122	1592.86	226.184
Water	8.07131	1730.63	233.426

Antoine equation $\log P_i^{sat} = A - B / (T + C)$ with P_i^{sat} in torr and T in °C.

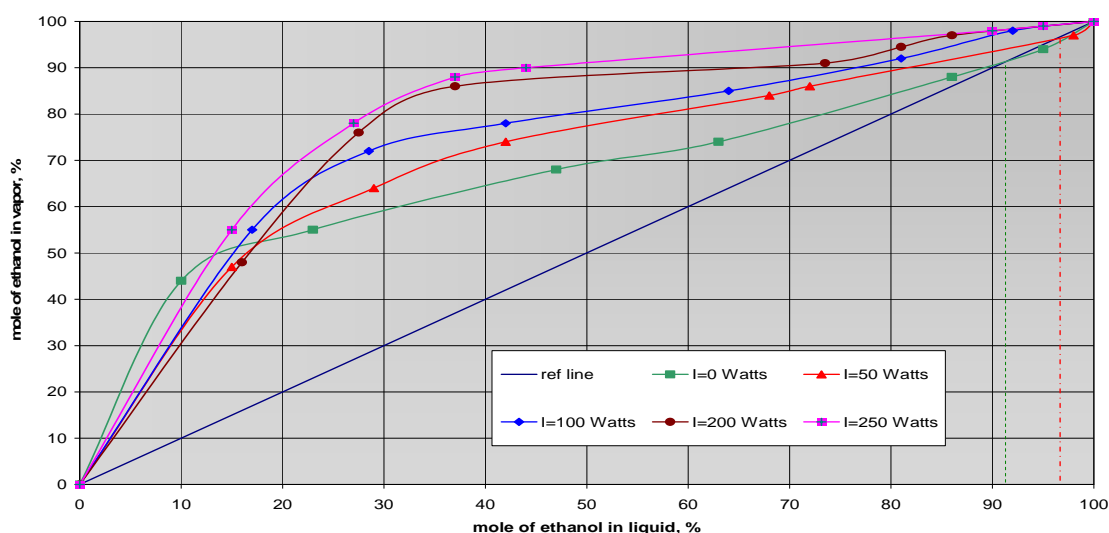
3.2 VLE study with sonication

The VLE data for ethanol-water with different ultrasonic power are shown in Figure 2. As can be seen from

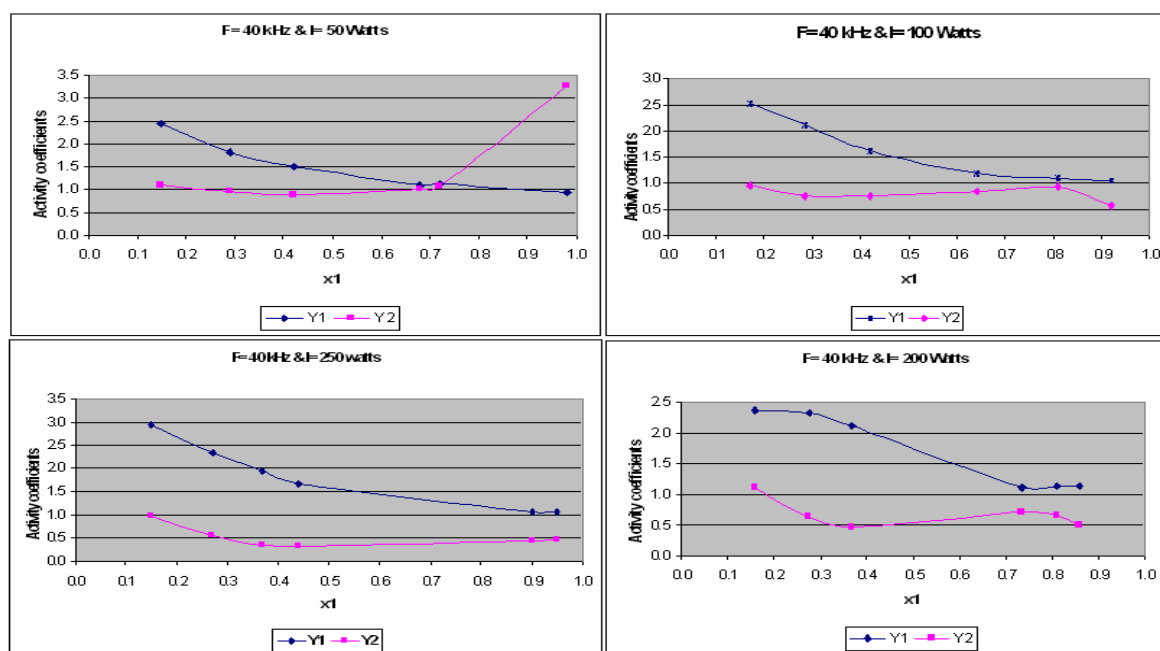
Figure 2, the VLE data of ethanol-water is altered with the presence of ultrasonic wave. In addition, the azeotropic point is shifted upward and then totally eliminated with the

increasing of ultrasonic power. For experimental study without the application of ultrasonic wave, the equilibrium curve crosses the reference line (45° diagonal line) at 92% moles of ethanol. Therefore, at this point, the ethanol-water mixture cannot be separated further since the vapor and liquid have identical compositions. When ultrasonic wave was applied at 50 Watts the azeotropic point was shifted to

96% of ethanol. However, the azeotropic point was totally eliminated when 100 Watts, 200 Watts and 250 Watts of ultrasonic wave was applied. Azeotropic compositions of ethanol-water system obtained during experimental study are shown in Table 4. The relation between azeotropic compositions and ultrasonic power are shown in Figure 3.



(a)



(b)

Figure 2: VLE data for ethanol–water system at atmospheric pressure with different ultrasonic powers: (a) x-y diagram (b) activity coefficients

Table 4: Azeotropic composition of ethanol-water

Ultrasonic Power (Watts)	Azeotropic Composition (% mole ethanol)
Without Ultrasonic, 0	92
50	96
100	100
200	100
250	100

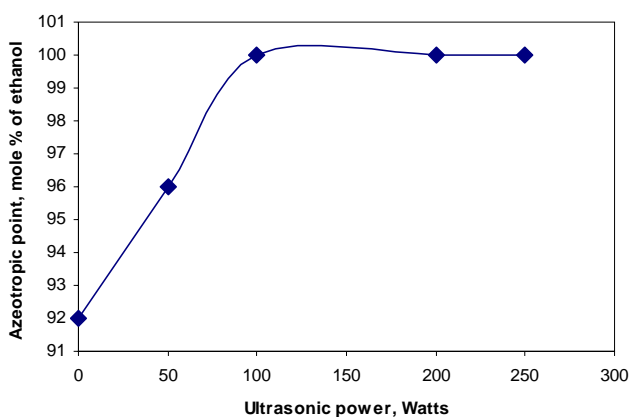


Figure 3: Azeotropic compositions as a function of ultrasonic power

Figure 3 shows the azeotropic point of ethanol-water system as a function of ultrasonic power. The azeotropic compositions are shifted and eliminate with the increase of ultrasonic power. This situation is due to the alteration of VLE data. Cavitation phenomenon which occurs during transmission of ultrasonic wave in liquid medium plays the important role in alteration of the VLE data. Through cavitation, the ultrasonic wave will submit the liquid mixtures to high acoustic pressure, producing a sudden increase and decrease of shear pressure within the liquid, together with local shock heating. This local shock heating will supply an additional heat to the system, increase turbulence and at the same time increase vaporization and mass transfer rate between liquid and vapor phases [17].

Due to differences in the physical and chemical properties of the individual component in the mixture, the amount of ultrasound energy being adsorbed by the components in the azeotropic mixture would vary accordingly. The volatile component in the liquid mixture, where the vapor pressure are lower, offers the most likeable spot of cavitation microbubbles occurrence [18]. The microbubble contains the vapor of the volatile component. Due to this situation the vapors released during the bubbles collapse will increase the concentration of the volatile component in vapor phase. Increasing amount of volatile component in vapor phase will increase the relative volatility of component in mixtures. The relative volatility of ethanol-

water system with different ultrasonic powers is tabulated in Table 5 and shown in Figure 4.

Table 5: Relative volatilities of ethanol (1)-water (2)

Ultrasonic power (Watts)	α_{12} (average)
0	2.8749
50	3.1379
100	4.5767
200	6.0989
250	8.518

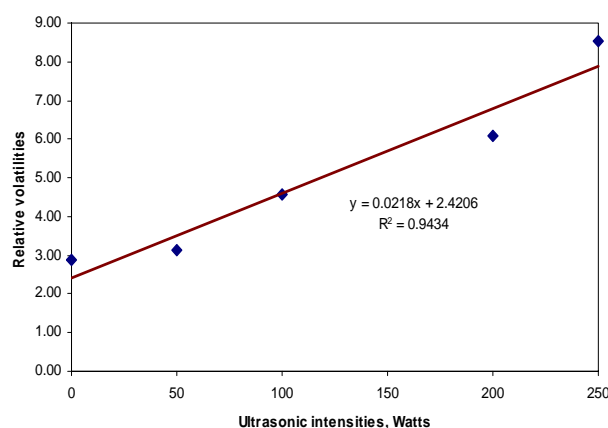


Figure 4: Relative volatilities of ethanol-water system against ultrasonic powers

Relative volatility plays the important role to ensure the successful separation of mixtures in distillation column. Relative volatility is a convenient index of relative ease or difficulty of separating components by distillation [19]. Separation of two components by distillation is impossible when the relative volatility is equal to 1.0 [20]. As shown in Figure 4, the relative volatilities are increase respectively with the increasing of ultrasonic power. The highest relative volatility obtained when the 250 Watts ultrasonic power was applied. The increasing in relative volatility will make the separation become easier. This phenomenon explains the effectiveness and practicability of this method in separating the ethanol-water mixtures.

4. Conclusions

Vapor-liquid equilibrium relation of binary system containing ethanol and water was measured at atmospheric pressure using equilibrium still immersed in the ultrasonic bath. The applications of ultrasonic wave enhanced the separation of ethanol-water and eliminate the azeotropic point by shifting the azeotropic point and at the same time altering the VLE of the system. The relative volatility of components

was increased respectively with ultrasonic power, making it possible to secure high purity product.

Acknowledgements

A special thanks is dedicated to every person involve in this research work. Special thanks are also dedicated to Universiti Teknologi Malaysia for their graduate scholarship to Siti Kholijah Abdul Mudalip.

List of symbols

A, B, C	Antoine constant in Table 3
P	pressure
P_i^{sat}	pure component vapor pressure
T	temperature (°C)
x_i	liquid phase mole fraction
y_i	vapor phase mole fraction

Greek letters

α_{ij}	relative volatility of component i and j
γ_i	liquid activity coefficient
η_D	refractive index

Subscripts

i, j, 1,2	component identification
-----------	--------------------------

References

- [1] Geankoplis, C. J. 1993. *Transport Processes and Unit Operations*. 3rd. ed. Englewood Cliffs, N. J.: Prentice-Hall.
- [2] Banat, F., Al-Asheh, S. and Simandl, J. 2002. Effect of dissolved organic salts on the isothermal vapor-liquid equilibrium of propionic acid-water mixture. *Journal of Chemical Engineering and Processing*, 41: 793-798.
- [3] Pinto, R.T.P., Wolf-Maciel, M.R. and Lintomen, L. 2000. Saline extractive distillation process for ethanol purification. *Journal of Computers and Chemical Engineering*, 24(2-7): 1689-1694.
- [4] Donis, I.R., Esquijarosa, J.A., Gerbaud, V., Fondevila, E.P., Joulia, X. 2004. Separation of n-Hexane-Ethyl Acetate mixtures by azeotropic batch distillation with Heterogeneous Entrainers. *Chemical Engineering and Processing*, 44 (1): 131-137.
- [5] Wolf Maciel, M.R and Brito, R.P. 1995. Evaluation of the dynamic behavior of an extractive distillation column for dehydration of aqueous ethanol mixtures. *Computers & Chemical Engineering*, 19(1): 405-40.
- [6] Lei, Z., Wang, H., Zhou, R., and Duan, Z. 2002. Influence of salt added to solvent on extractive distillation. *Chemical Engineering Journal*, 87(2): 149-156.
- [7] Ladisch, M.R. and Dyek, K. 1979. Dehydration of ethanol: a new approach gives positive energy balance. *Science*, 205: 898.
- [8] Lee, F.M. and Pahl, R.H. 1985. Solvent screening and conceptual extractive distillation process to produce anhydrous ethanol from fermentation broth. *Industrial Engineering Chemical Process Design and Development*, 24: 168.
- [9] Khalfaoui, B., Menidi, A.H., Borja, R. 1997. Thermodynamic properties of water + normal alcohols and vapor-liquid equilibria for binary system of methanol or 2-propanol with water. *Fluid Phase Equilibria*, 127: 181-190.
- [10] Fairbanks, H.V. 2001. Drying powdered coal with the aid of ultrasound. *Powder Technology*, 40(1-3): 257-264.
- [11] Mason, T.J. and Lorimer, J. (1996). Ultrasonic intensification of chemical processing and related operation- A Review. *Transactions of The Institute of Chemical Engineers*. 74 A: 511-516.
- [12] Povey, M.J.W. 1998. Ultrasonics of food. *Contemporary Physics*. 39(6): 467-478.
- [13] Gomis, V., Francisco, R. and Juan Carlos, A. 2000. The application of ultrasound in the determination of isobaric vapor-liquid-liquid equilibrium data. *Fluid Phase Equilibrium*, 172(2): 245-259.
- [14] Gomis, V., Font, A., Pedraza, R., and Saquete, M.D. 2005. Isobaric vapor-liquid and vapor-liquid-liquid-equilibrium data for the system water+ethanol+cyclohexane. *Fluid Phase Equilibrium*, 235(1): 7-10.
- [15] Alberto, A., Rodil, E., and Soto, A. 2005. Isobaric vapor-liquid equilibria for systems composed by 2-ethoxy-2-methylbutane, methanol or ethanol and water at 101.32 kPa. *Fluid Phase Equilibria*, 233(1): 9-18.
- [16] Gmehling and Onken 1977. Vapor-liquid equilibrium data collection, DECHEMA Chemistry Data Series, Vol 1, parts (1-10): Frankfurt.
- [17] Mason, T.J., 1991. *Practical Sonochemistry: user's guide to application in chemistry and chemical engineering*, Ellis Horwood Limited: New York.
- [18] Zhou, D.W., Liu, D.Y., Hu, X.G., and Ma, C.F. 2002. Effect of acoustic cavitation on boiling heat transfer. *Experimental Thermal and Fluid Science*, 26(8): 931-938.
- [19] Van Winkle, M. 1967. *Distillation*. New York: McGraw-Hill Book of Company.
- [20] Shinkey, F. G. 1977. *Distillation control for productivity and energy conservation*. New York: McGrawHill Book Company.

The Use of Surfactant Modified Zeolite Y as a Sorbent for Arsenic Species in Aqueous Solution

Nik Ahmad Nizam¹ and Alias Mohd Yusof^{1*}

¹ Department of Chemistry, Faculty of Science
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5534500, Fax: +60-7-5566162, E-mail: g_doxnick@yahoo.com

Abstract

The removal of the trivalent arsenic (As(III) or arsenite) and pentavalent arsenic (As(V) or arsenate) from aqueous solution by surfactant modified zeolite Y (SMZY) was studied. The SMZY was prepared by attaching the surface of the zeolite NaY with cationic surfactant, hexadecyltrimethyl ammonium (HDTMA) which exhibited 100% external cation exchange capacity (ECEC) of the zeolite NaY. While the unmodified zeolite Y had no or a little affinity for As(III) (neutral species) and As(V) (anion species), the surfactant modified zeolite Y with HDTMA showed significant removal of both arsenic species from aqueous solutions. The maximum adsorption of HDTMA onto zeolite (1111.1 mmol/kg) was nearly twice the ECEC value of the zeolite NaY (671.4 mmol/kg) revealing that the HDTMA molecule with concentrations above critical micelle concentration (CMC) creates an organic rich layer on the zeolite surface and the charge on the surface is reversed from negative to positive. Thus, the positive charge from HDTMA will be exchanged by arsenate (As(V)) anion while the organic partitioning will absorb the arsenite (As(III)) species.

Keywords:

As(III); As(V); Surfactant Modified Zeolite; Zeolite NaY

1. Introduction

The element arsenic which is prominently carcinogen, can form anion as arsenate (As(V)) and nonionized species as arsenite (As(III)) in the aqueous phase, therefore, the materials that can remove this kind of toxic metals simultaneously in the water body are of great importance with regards to the water purification or wastewater treatment. To achieve this result, the composite ion exchanger has to be developed, for which it shall contain the properties of anion exchanger and neutral partitioning at the same time. However, a typical zeolite cannot remove or sorb the anion and neutral species as its surface is in the anionic charges form. By treating the zeolite with a cationic surfactant, an organic covering is created on the external surfaces and the charge is reversed to positive.

Therefore, the modification surface of zeolite by cationic surfactant (HDTMA) can be done according to the successful clinoptilolite (naturally occurring zeolite) modified with HDTMA by Li and Bowman [1] and they called it surfactant-modified zeolite (SMZ). Laboratory batch and column tests demonstrate that SMZ can simultaneously remove multiple types of contaminants from water, including inorganic anions such as chromate and

hydrophobic organics such as chlorinated solvents and fuel components [2]. Zeolite NaY resembles natural zeolite minerals that have permanent negative charges on their surface and large cation exchange capacity (CEC) which enable them to be modified by cationic surfactant to enhance their sorption of neutral and anionic contaminants in water.

The surfactant that is commonly employed to be attached on the zeolite surface in the previous studies is HDTMA, the quaternary amine hexadecyltrimethyl ammonium cation which is a long chain cationic surfactant that possesses a permanent positive charge. The HDTMA is in the grouped under cationic surfactants where they have positively charged hydrophilic head group which is generally amine, attached to a hydrophobic tail of hydrocarbon moiety. The HDTMA structure consists of permanently charged trimethyl ammonium head group attached to a 16-carbon chain. Since the uses of this surfactant are mainly as hair conditioner, mouthwash and fabric softeners, it is assumed that low levels of HDTMA will not be harmful to the environment.

The objectives of this study is to modify the zeolite Y surface by treating it with cationic surfactant, HDTMA in an amount equal to 100% of the zeolite external cation exchange capacity. The surfactant modified zeolite Y was then used to study the effectiveness of its capacity towards the removal of arsenate and arsenite from aqueous solutions.

*Corresponding Author. E-mail: Alias@kimia.fs.utm.my,
Tel: +60-7-5534500, Fax: +60-7-5566162

2. Methods and Materials

2.1. Raw Materials and Chemicals

The zeolites used is the zeolite NaY synthesized from rice husk ash (silica source) via seeding technique and the commercial zeolite NaY supplied by Zeolyst International (CBV 100) for comparison. The physicochemical properties of the zeolites related to the use of zeolites as a sorbent is given in table 1.

Table 1. The physicochemical properties of both zeolites

Physicochemical properties	Synthesized zeolite NaY	Commercial Zeolite NaY
Unit cell, a_0 (Å)	24.759	24.7
$^a\text{SiO}_2/\text{Al}_2\text{O}_3$	3.871	4.553
Surface area (m^2/g)	506.6	484.9
Cation exchange capacity, CEC (meq/g)	3.15	2.545
External cation exchange capacity, ECEC (meq/g)	0.6714	0.533

Notes: a: Calculated from unit cell

The cationic surfactant used to modify the zeolite surface was hexadecyltrimethyl ammonium bromide (HDTMA-Br) powder supplied by MERCK-Schuchardt. The stock solution of arsenite (1000 mg/L) was prepared by dissolving As_2O_3 (1.32 g) in solution containing 4 g of NaOH pellet in 100 mL and after the dissolution was completed, the mixture was then added with 20 mL concentrated HNO_3 and finally diluted to the 1000 mL. The stock solution of arsenate was prepared by dissolving Na_2HAsO_4 (0.0416 g) in a solution containing 0.2 g NaOH in 10 mL and finally diluted to the 50 mL with distilled water.

2.2. Adsorption of HDTMA on Zeolite NaY

The synthesized zeolite NaY (0.5 g) was weighed accurately and added with the HDTMA solutions to yield cation dosages in the range from 25 to 250% of zeolite ECEC. The suspensions of zeolites and HDTMA cation were shaken for 24 hours at a constant agitation rate and the ambient temperature. For good separation of solid and liquid, the tube was then centrifuged (5000 rpm, 15 minutes) and subsequently the supernatant was decanted. The liquid supernatant was then analyzed for the determination of the concentration of the remaining HDTMA in the solution using the total organic carbon analyzer (TOC-Ve, Shimadzu, Japan). About 20 μL of the supernatant was taken using a syringe and was injected into the injection port of the TOC instrument. The concentration of the total organic carbon was determined automatically by the instrument. Assuming complete conversion of organic cations to CO_2 during TOC analysis, the TOC content of samples were converted to the equilibrium concentration of HDTMA (mmol/L).

2.3. Preparation of Surfactant Modified Zeolite Y (SMZY)

HDTMA solution was added in an amount equal to 100% of external cation exchange capacity (ECEC) of the zeolite. An appropriate weight of dried zeolite was mixed with 900 mL of aqueous solutions containing a suitable weight of HDTMA cation. After the aqueous solutions were mixed with zeolites, the mixture was stirred for 5 days at room temperature. The mixture was then filtered and the solid sample was dried at 60 °C overnight. The resultants SMZY can be readily used for the adsorption study.

2.4. Adsorption of Arsenic

About 0.1 g of samples (SMZY and both unmodified zeolite Y) were weighed and placed in the 50 mL centrifuge tube containing the single component of As(III) (20 mg/L) or As(V) (pH 8, 20 mg/L). The centrifuge tube was then shaken for 24 hours using an orbital shaker at constant agitation rate (120 rpm). The solid sample and solution was separated and the filtrate solution was analyzed for the concentrations of the remaining As(V) or As(III) in the solution by Flame Atomic Absorption Spectroscopy (FAAS) (Perkin Elmer, model AAnalyst 400). The adsorption capacities of both species were stated as the percentage of the adsorption.

3. Results and Discussion

3.1. Maximum Adsorption of HDTMA on Zeolite NaY

Sorption of cationic surfactants from solution onto solid surfaces can be described by the Langmuir isotherm. The Langmuir equation is based on a kinetic approach and assumes a uniform surface, a single layer adsorbed material constant temperature. The Langmuir equation is given as:

$$q_e = \frac{bQ_o C_e}{1 + bC_e} \quad (1)$$

where:

- q_e : amount adsorbed at equilibrium (mmol/kg)
- C_e : equilibrium concentration (mmol/L)
- b : Langmuir constant related to the affinity of the binding site.
- Q_o : maximum adsorption at monolayer coverage (mmol/kg)

The nonlinear form (equation 1) can be evaluated by transforming it to the linear equation:

$$\frac{m}{x} = \frac{1}{q_e} = \frac{1}{Q_o} + \frac{1}{bQ_o C_e} \quad (2)$$

When $\frac{m}{x}$ or $\frac{1}{q_e}$ against $\frac{1}{C_e}$ is plotted, a straight line is obtained with a slope $\frac{1}{bQ_o}$ and intercept-Y is $\frac{1}{Q_o}$.

The isotherm of the adsorption of HDTMA on synthesized zeolite NaY is given in Figure 1.

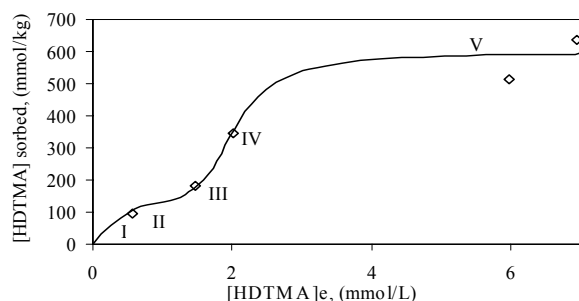


Figure 1. The sorption isotherm plotted of HDTMA onto the synthesized zeolite Y.

As shown in this figure, the shape of the Langmuir plot for the sorption of HDTMA onto the synthesized zeolite Y was identical to the theoretical adsorption of the surfactant on the solid having highly hydrophilic solid [3]. The adsorption of a surfactant from solution onto a surface depends on the concentration of surfactant. At very low concentrations (I and II), there is no orientation and the molecules lie flat on the surface. As the concentration increases (III), the number of surfactant molecules on the surface increases, there is not enough space for them to lie flat and so they begin to orient, the orientation depending upon the nature of the hydrophilic group and the surface. At concentration IV, the number of surfactant molecules available is now sufficient to form a unimolecular layer. This particular concentration is important and known as the critical micelle concentration (CMC). At concentration V (above IV), since the zeolite has hydrophilic surface, more than one layer of surfactant molecules can form ordered structures on the surface of the zeolite. In addition, the surfactant molecules in the solution will form an ordered structure known as micelle as long as the concentration is above the CMC. Data set described by the Langmuir isotherm with coefficient of determination (r^2) around 0.97 (Table 2) when the data were plotted according to equation 2.

Table 2. Fitted Langmuir parameters for sorption of HDTMA by synthesized zeolite Y

Q_o (mmol/kg)	b (1/kg)	r^2
1111,1	0,1607	0,9754

The HDTMA molecule is sorbed (1111.1 mmol/kg) essentially quantitatively up to nearly twice the ECEC value of the synthesized zeolite Y (671.4 mmol/kg). Sullivan et al. [4] also found the same results when they used clinoptilolite, the naturally occurring zeolite tailoring with HDTMA. Thus proved that the HDTMA molecule with concentration above CMC creates an organic-rich

layer on the zeolite surface and the charge on the surface is reversed from negative to positive. The positive charge on the outward-pointing HDTMA head groups is balanced by anions from solution, forming an electrical double layer. The schematic diagram for this theory can be seen in Figure 2. The anion properties in the surface of surfactant modified zeolite allow them to exchange the counter ions by the other anions.

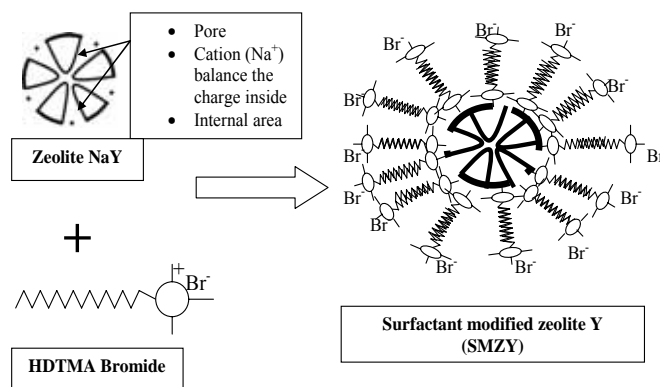


Figure 2. The schematic diagram of the theoretical HDTMA formation on the zeolite Y structure.

3.2. Adsorption of Arsenic.

The SMZY and unmodified zeolite Y were both put in contact with the single solute in single concentration of arsenate and arsenite solution. The sorbed concentration of arsenite (As(III)) and arsenate (As(V)) was calculated from the difference between the final concentration that is after the adsorption and the initial concentration of both arsenic species. The results from these experiments are present in Table 3 and illustrated in Figure 3.

Table 3. The values of the adsorption of As(III) and As(V) by SMZY and respective parent zeolite Y.

Samples	[As(III)] _{sorbed} , mg/g	[As(V)] _{sorbed} , mg/g
Synthesized zeolite NaY	0.127	0.047
SMZY-Synthesized	0.184	0.923
Commercial Zeolite NaY	0.080	0.011
SMZY-Commercial	0.161	0.512

As shown in Figure 3, it is clearly observed that the SMZY adsorbs more As(III) than the unmodified zeolite Y but not as significant as the adsorption of As(V) by SMZY. Since the species of As(III) forms neutral species in water, the adsorption of As(III) on SMZY is due to the partition of these species into the organic partitioning created from the bilayer of HDTMA onto the surface of the zeolite Y at HDTMA loading greater than the external cation exchange capacity (ECEC). The organic partitioning of the HDTMA can be seen in Figure 2. Furthermore, the zeolite itself can

adsorb the neutral form of As(III) through the adsorption on the surface of the zeolite since both zeolites have high specific surface area. Besides that, As(III) is removed by unmodified synthesized zeolite Y than the commercial one indicating that the structure, including SiO₂/Al₂O₃ ratio and the specific surface area as well as CEC, play an important role in the arsenic removal process. As a comparison from previous literature, the sorption of As(III) by zeolite Y was more than the natural zeolite, clinoptilolite (0.014 mg/g) [5], due to the higher CEC of the synthetic zeolite Y.

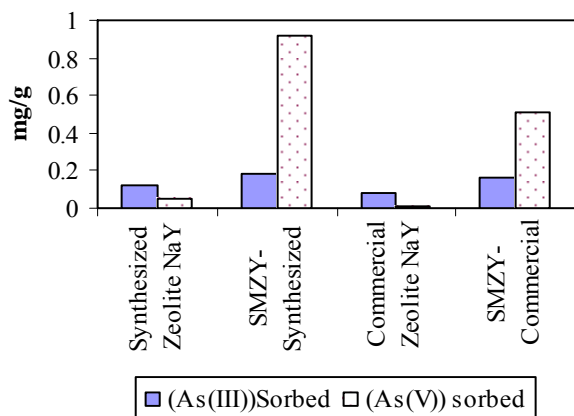


Figure 3. The adsorption of As(III) and As(V) by modified and unmodified zeolite Y

Figure 3 also demonstrates that both unmodified zeolite NaY have little affinity for As(V), in contrast, the SMZY shows significant sorption of As(V) from aqueous solutions. The ability of the SMZY to sorb As(V) is due to the anion exchange at the positive charge brought about by the reversed charge resulting from the HDTMA double layer onto the zeolite surface. The mechanism of the uptake As(V) by SMZY can be illustrated in figure 4. Because the As(V) exists in water as the anion form and analogous with the Cr(VI) species, the SMZY is able to sorb As(V) by the anion exchange process. On the contrary, the unmodified zeolite NaY having a net negative charge in the structure tends to repulse the anion form of As(V) species. The SMZY prepared from the synthesized zeolite Y tend to sorb more As(V) than the SMZY from the commercial one. The noteworthy difference of the As(V) sorption by SMZY from the synthesized one compared to the SMZY from the commercial one is due to the properties of the parent synthesized zeolite Y that has higher CEC, higher surface area and lower ratio of silica per alumina than the commercial zeolite Y. These properties strongly affect the adsorption capacity of the As(V) by SMZY.

As illustrated in the schematic diagram in Figure 4, the sorption of arsenate anion by the cationic surfactant modified zeolite Y (SMZY) is contributed by surface anion exchange on the positively charged outermost surface of the zeolite. The positively charged outermost surface of SMZY was created by surfactant bilayer sorption when the amount of surfactant loading exceeded the external cation exchange capacity (ECEC) of the zeolite. In addition, the sorption of arsenate anion involved on the replacement of weakly held

counterions by strongly held counterions will eventually lead to the formation of a surface-anion complex. The formation of the SMZY-arsenate is insoluble, thus it will remain as solid particles in the solution that can be separated later by a simple filtration technique.

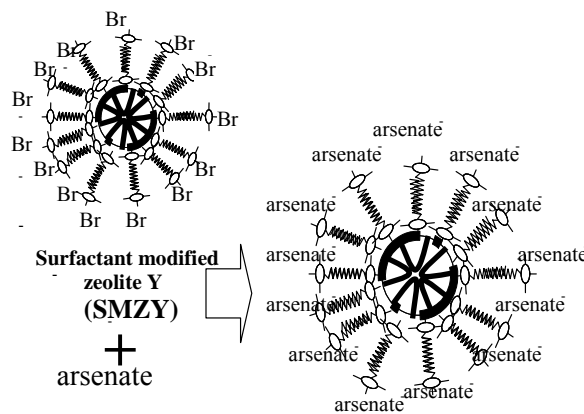


Figure 4. The schematic diagram for the mechanism of As(V) sorption by SMZY

4. Conclusions

While the unmodified zeolite NaY have little affinity towards the anion and the neutral species, the zeolite NaY that have been treated with cationic surfactant (HDTMA) in an amount equal to 100% of its external cation exchange capacity (ECEC) improved the efficiency in removing these species. The surfactant modified zeolite Y (SMZY) was useful and efficient to remove arsenate in the form of anion and the arsenite in the neutral form. The HDTMA which created the bilayer formation through tail-tail interaction via columbic forces on the surface of zeolite Y resulted by the reverse charge from negative to positive charge was found to be the factors governing the inorganic anion and neutral species sorbed by the SMZY. Therefore, the zeolite treated with HDTMA with fully satisfied ECEC zeolite can be used an alternative sorbents in removing the As(III) and As(V) species from water.

Acknowledgments

We would like to express our gratitude to MOSTE (IRPA 09-02-0057 SR0005/09-9) for the financial support and all the members of Chemistry Department, Faculty of Science, Universiti Teknologi Malaysia in assisting this research.

References

- [1] Li, Z. and Bowman, R. S. 1997. Counterion effects on the sorption of cationic surfactant and chromate on natural Clinoptilolite. *Environ. Sci. Technol.* 31: 2407-

- 2412.
- [2] Li, Z., Anghel, I. and Bowman, R. S. 1998. Oxyanion sorption by surfactant modified zeolite. *J. Dispersion Sci. Technol.* 19: 843-857.
- [3] Porter, M. R. ed. 1994. *Handbook of surfactants*. Glasgow: Blakie academic and professional.
- [4] Sullivan, E. J., Carey, J. and Bowman, R. S. 1998. Thermodynamics of cationic surfactant sorption onto natural clinoptilolite. *J. Colloid Interface Sci.* 206: 369-380.
- [5] Elizalde-Gonzalez, M. P., Mattusch, J., Einicke, W. D. and Wennrich, R. 2001. Sorption on natural solids for arsenic removal. *Chem. Eng. J.* 81: 187-195.

Immobilization of Glucose Oxidase and Ferrocene Redox Polymer in cross-linked poly (vinyl alcohol) with bovine serum albumin as Protein Stabilizer

Norhana Jusoh, A Abdul Aziz

*Faculty of Chemical and Natural Resources Engineering,
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia.
Tel: +60-7-5535621, Fax: +60-7-5581463 Email: azila@fkkksa.utm.my*

Abstract

A method of tethering a mediator to an enzymatic membrane was studied. Glucose oxidase (GOD) and ferrocene redox polymer were immobilized in cross-linked poly (vinyl alcohol) (CLPVA) with bovine serum albumin (BSA) added as a protein stabilizer. Redox hydrogel polyallylamine ferrocene was prepared by cross-linking polyallylamine hydrochloride with glutaraldehyde and attaching the ferrocene covalently. The biosensor response to glucose was evaluated amperometrically at 0.363V. High BSA concentration resulted in improved current response but lower K_m^{app} . CLPVA, which had been proven to be an excellent retainer of GOD could not retain both GOD and ferrocene redox polymer effectively, thus affecting the stability of the enzymatic membrane.

Keywords

Biosensor, Glucose Oxidase, Redox Polymer, Cross-linked PVA, Amperometric.

1. Introduction

A biosensor is a sensor that is based on the use of a biological material for its sensing function. The bio-component specifically reacts or interacts with the analyte of interest resulting in a detectable chemical or physical change. Three general strategies are used for the electrochemical sensing of glucose. All methods employ glucose oxidase (GOD), an enzyme that catalyzes the oxidation of glucose to gluconic acid with the production of hydrogen peroxide. The first detection scheme measures oxygen consumption; the second measures the concentration of hydrogen peroxide produced by the enzyme reaction; and the third uses a diffusible or immobilized mediator to transfer the electrons from glucose oxidase to the electrode.

The use of an artificial electron acceptor or mediator to replace the natural acceptor oxygen in the oxidation of glucose by glucose oxidase is a preferable approach that has been explored to overcome tissue oxygen dependence. In addition, the oxidation of the reduced mediator occurs at low potential thus reducing the sensitivity of the sensor to interfering substances such as uric acid, ascorbic acid and acetaminophen. These substances will break down electrochemically and thus give interfering signals. In addition, mediated biosensors offer other advantages such as increased linear response and perhaps an extended biosensor lifetime, because hydrogen peroxide, which can contribute to the deactivation of the enzyme, is not being generated, [1].

After years of research, progress is finally being made toward implantable continuous glucose monitors. Most implantable glucose sensors are amperometric enzymatic biosensors. The first report on an implantable amperometric ferrocene-modified glucose sensor was in 1986 [2]. However, the initial promise exhibited by mediator based

glucose sensors for in vivo applications, has failed to materialize. The main problem remains the limited long-time-use stability of mediated glucose sensors, which has been attributed to the leaching of the mediator. In addition, the loss of mediator is a particularly important issue for implantable sensors because of the inherent toxic effect of the mediators used. Therefore, in order to develop a stable implantable mediated glucose sensor, a suitable immobilization method should be investigated to avoid the leaking of mediator as well as the enzyme.

A promising strategy in biosensor design is the immobilization of both enzyme and mediator, which generally require polymeric material. Different redox polymers have been used as mediators including osmium polymers [3,4], ferrocene polymers [5] and Nafion-N-methyl phenazinium [6]. The advantages of using the redox polymer are several with the main advantage is more stable biosensors since leaking of mediator from the electrode is minimized and higher and faster responses are observed due to proximity between the enzyme and the mediator [7].

In this study, the enzyme and the ferrocene redox polymer (Fc) were immobilized in cross-linked PVA (CLPVA). Bovine serum albumin (BSA) was added to stabilize the enzyme. CLPVA was applied as a solid support due to the ability to form very homogenous films with very high quality. Redox hydrogel polyallylamine ferrocene was prepared by crosslinking polyallylamine hydrochloride with glutaraldehyde and attaching the ferrocene covalently [8]. Amino group of cross-linked polyallylamine and carboxyl group of ferrocene carboxylic acid were activated using carbodiimide reagents. Larger response currents were expected with a polycationic redox hydrogel such as derivatized polyallylamine [9]. The effect of BSA loading also was investigated.

2. Approach and methods

2.1 Chemicals

Glucose oxidase (E.C. 1.1.3.4) from *Aspergillus niger* were purchased from Sigma (England). Ferrocene carboxylic acid (97%) was purchased from Aldrich (Germany). Cystamine dihydrochloride (98%) were purchased from Aldrich (China). Peroxidase horseradish (E.C. 1.11.1.7, type VI from Horseradish), glucose (corn sugar, 99.5%), poly (allylamine hydrochloride) (Average MW CA:70 000), HEPES (99.5%, pH 6.8-8.2), bovine serum albumin (BSA), polyvinyl alcohol (PVA, Average MW 70 000-100 000), glutaraldehyde and *N*-cyclohexyl-*N'*-(2-morpholinoethyl) carbodiimide metho-*p*-toluenesulfonate were purchased from Sigma (USA). Tetramethylorthosilicate (TMOS), kalium di-hydrogen phosphate, di-kalium hydrogen phosphate, kalium chloride, acetic acid, methanol, sulfuric acid and hydrochloric acid were purchased from Merck (Germany). All chemicals were used as received.

2.2 Synthesis of poly(allylamine) ferrocene (PAA-Fc).

Preparation of ferrocene-containing redox polymer was done according to a previous work [8]. 581 mg of polyallylamine hydrochloride and 5 mL of 20% glutaraldehyde solution were dissolved in a HEPES buffer (50mM, pH6.8) to a total volume of 25mL in a beaker and was then left to gelate. The cross linked gel was crushed through a sieve and freeze dried. 60mg of this polymer was suspended in 50mL of HEPES buffer (50mM, pH6.8) containing 115mg of ferrocene carboxylic acid.

Water soluble carbodiimide was added drop wise during the first hour. The reaction was allowed to proceed for 4 days. Small particles of the ferrocene modified polyallylamine hydrogel were rinsed with a phosphate buffer solution. These particles were enclosed in dialysis tubes containing phosphate buffer. The dialysis was carried out for 3 days.

2.3 Cross linking with PVA and BSA addition

5% PVA stock solution was prepared by dissolving PVA in water and heating the solution to 80–90 °C under stirring for about 30 minutes. Then, the 5% PVA stock solution was mixed with 10% acetic acid as a buffer, 50% methanol as a quencher, and 10% sulphuric acid as a catalyst in the volume ratio of 5: 3: 2: 1. Appropriate amount of 2% glutaraldehyde was added to the solution in order to obtain a cross-linking ratio of 0.06. Cross-linking ratio is defined as the ratio of the moles of glutaraldehyde per moles of PVA repeat unit. Then, polyallylamine ferrocene, BSA and GOD were added to the CLPVA solution and an aliquot of the mixture was pipetted on a glass slide and air-dried for 20 minutes. Then, it was covered with another glass slide and the two glass slides were clamped together and left for 24hr at 4°C. The membranes obtained were swollen in phosphate buffer at 4°C.

2.4 Enzyme and mediator leakage detection

Leakage of enzyme was measured colorimetrically. The chromogen solution was prepared by diluting 0.1 mL of 1% O-dianisidine in 12 mL of 0.1 M phosphate buffer, pH 6.7. Then, 150µL of 18% aqueous glucose solution and 50µL of 200µg/mL peroxidase solution were added to 1.25 mL of the chromogen solution. The mixture was then placed in a water bath at 25°C for temperature equilibration. Then, 50µL of the washing solution was added to the mixture. The reaction was allowed to proceed for 5 minutes before 100 µL of 4 M HCL was added to stop the reaction. The amount of colour formed was measured by reading the absorbance value at 450nm [10].

Leakage of ferrocene derivatives mediator was measured electrochemically. The washing solution was subjected to cyclic potentials from 600mV to -100mV with scan rate 10mVs⁻¹. The concentration of the mediator was determined using a calibration curve.

2.5 Electrochemical measurement

Electrochemical measurements were carried out using a potentiostat with a three-electrode configuration (Metrohm µAutolab Type 111). The working electrode (WE) used was a platinum electrode. A platinum auxiliary electrode was used as the counter electrode (CE). An Ag/AgCl/ KCl was employed as the reference electrode (RE).

Before use, the enzyme electrode was rinsed with doubly distilled water, and immersed in 0.1M phosphate buffer (pH7.0) until a stable electrochemical response was produced. Glucose stock solutions was allowed to mutarotate at room temperature overnight before use. All solutions were deoxygenated before each amperometric run. For kinetics, response time and stability studies, the amperometric experiments were run at 363 mV vs Ag/AgCl. All experiments were performed at a temperature of 25±1 °C and under deoxygenated condition, unless otherwise specified.

3. Results

3.1 Retention of enzyme and mediator in membranes

To investigate the ability of the membranes to retain GOD and ferrocene mediator, the washing solutions for the CLPVA-GOD/Fc membranes were assayed for any sign of enzyme activity and also leakage of the mediator. Figure 1(a) and 1(b) show the leaking profiles of GOD and ferrocene for the CLPVA-GOD/Fc membranes.

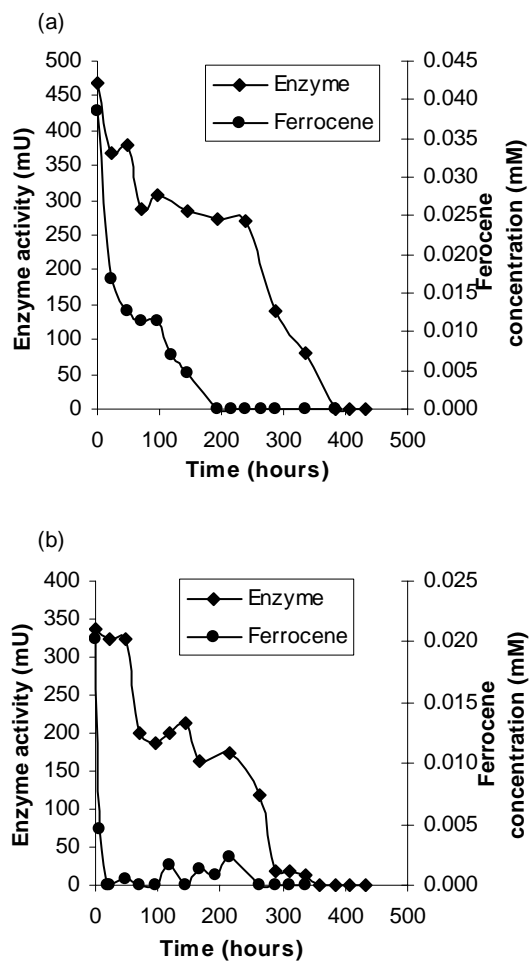


Figure 1. Leaking profile for CLPVA-GOD/Fc membrane with different GOD and BSA loading a) 1:1 b) 1:3 (weight ratio of GOD: BSA).

As shown in figure 1(a) and 1(b), the leaking of enzyme as well as mediator decreased with time. No sign of enzyme activity was observed in the washing solutions after 15 days for membranes with the weight ratio of 1:3 (GOD: BSA), which was 1 day earlier compared to membranes with the weight ratio of 1:1 (GOD: BSA). Meanwhile, leakage of ferrocene from membranes with the weight ratio 1:3 (GOD: BSA) stopped after 11 days, which was 2 days later than the membranes with the weight ratio of 1:1 (GOD: BSA).

3.2 Kinetics properties of the membranes

Current response vs glucose concentration curves for both CLPVA-GOD/Fc membranes are shown in Figure 2. The kinetic properties of the membranes were determined from the modified electrochemical Lineweaver-Burke plots (Figure 3).

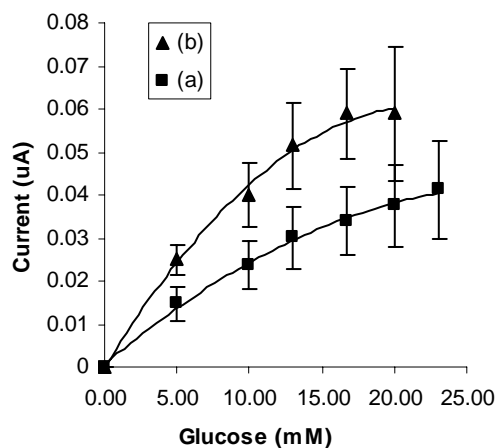


Figure 2. Current response vs glucose concentrations curves for CLPVA-GOD/Fc membranes with different GOD and BSA loading a) 1:1 b) 1:3 (weight ratio of GOD: BSA).

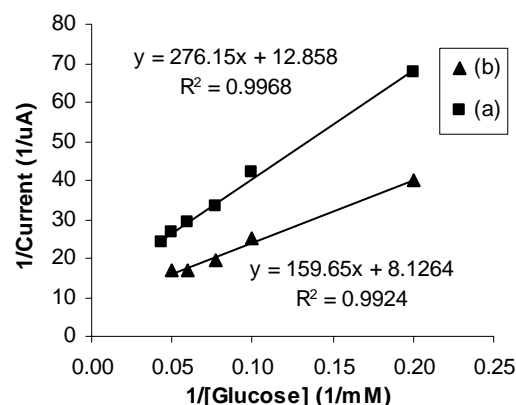


Figure 3. Double-reciprocal (Lineweaver Burke) plots of CLPVA-GOD/Fc membranes with different GOD and BSA loading a) 1:1 b) 1:3 (weight ratio of GOD: BSA in mg).

The apparent Michaelis-Menten constant, K_m^{app} for membranes with weight ratio (GOD: BSA) 1:1 and 1:3 were approximately, 21.48mM and 19.66mM, respectively. Meanwhile, the corresponding maximums current, I_{max} for both cases were 0.08uA and 0.12uA, respectively.

3.3 Stability of CLPVA-GOD/Fc membranes

Stability of CLPVA-GOD/Fc membranes was investigated to determine the shelf life of the sensors. The current outputs of the membranes to 5mM glucose at certain period were measured. Figure 4 shows the effect of storage time on stability of CLPVA-GOD/Fc membranes.

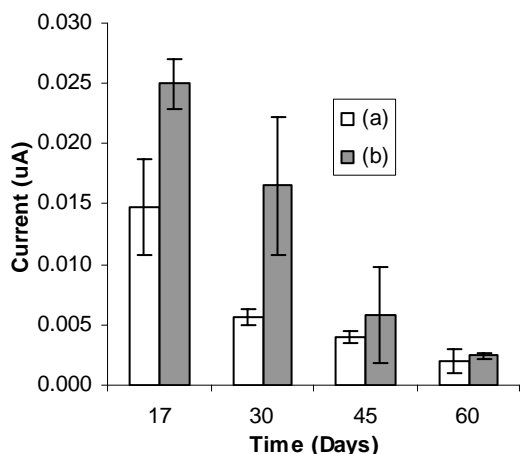


Figure 4. Stability of CLPVA-GOD/Fc membranes with different GOD and BSA loading a) 1:1 b) 1:3 (weight ratio of GOD: BSA).

As shown in figure 4, the membranes retained approximately only 38.87% and 66.00% of the initial current after 1 month, for membranes with weight ratio (GOD: BSA) 1:1 and 1:3, respectively. Then, after 2 month, only 3.5% and 9.7% of the initial current remained, respectively for both membranes.

4. Discussion

The retention of enzyme and mediator in the membranes were very poor although CLPVA was applied as a solid support. For both membranes, the leakage of ferrocene stopped earlier compared to the enzyme. However, the leaking of ferrocene should not have occurred since ferrocene was covalently attached to the polyallylamine hydrogel. The leaking might be due to high concentration of enzyme as well as ferrocene redox polymer that might have exceeded the immobilization capacity of the membranes. The excess enzymes and mediator were not immobilized within the solid support and leached out easily from the membrane.

The membranes with higher BSA gave higher current response towards glucose. BSA stabilized the enzymes, creating a 'biological like' environment. Albumin improves enzymatic activity because of better mass distribution of the various proteins without altering the mechanical properties of the membrane. BSA could also prevent the polymer matrix from over-swelling [8], which could extend the distance between the redox sites of the polymer. As the distance increased the electron transfer rate among neighbouring redox sites would decrease.

The apparent Michaelis-Menten constants, K_m^{app} were 21.48mM and 19.66mM respectively for membranes with weight ratio (GOD: BSA) of 1:1 and 1:3. These values were larger than the K_m^{app} of glucose oxidase in solution that has been reported to be approximately 12.43mM and 15.94mM at temperature 25°C and 30°C, respectively [11].

Generally, the K_m^{app} of an immobilized enzyme will be larger than that of the free enzyme in solution due to the effect of the diffusion of substrate to the active sites [10]. In this work, membranes with high loading of BSA had lower K_m^{app} . The low K_m^{app} suggested that the enzyme had a high affinity for the substrate [12].

As shown in figure 4, the stability of CLPVA-GOD/Fc membranes was not good. This could be due to the deterioration of the immobilized GOD or problems with the mediator. Brooks et al., however, reported that the loss of activity of ferrocene glucose sensors was more strongly influenced by the loss of enzyme by denaturation or detachment [13]. The addition of extra ferrocene to spent electrodes did not affect activity but the addition of more glucose oxidase rejuvenated the sensitivity to glucose. Thus, stability of CLPVA-GOD/Fc membranes could be improved if the immobilization process was more effective.

5. Conclusions

In this work, immobilization of glucose oxidase and ferrocene redox polymer in CLPVA with the addition of BSA has been done. A membrane with greater BSA content gave higher current response but had a smaller K_m^{app} , which might eventually decrease the detection limit of the biosensor. Extensive study must be done to improve the retention of enzyme and mediator as the CLPVA which had been shown to be an excellent retainer of GOD [10] was not able to retain both GOD and ferrocene redox polymer effectively. This would ultimately influence the stability of the membranes.

Acknowledgement

This work was supported financially by Intensification of Research in Priority Areas (IRPA), project no: 03-02-06-0092 EA001 and UTM- PTP scholarship.

References

- [1] Reynolds, E. R., Geise, J. R. and Yacynych, A. M. 1992. Optimization Performance Through Polymeric Material. In *Biosensor & Chemical Sensors: Edelman, P. G and Joseph Wang*, 106. USA: Library of Congress.
- [2] Claremont, D. J., Sambrook, I. E., Penton, C. and Pickup, J. C. 1986. Subcutaneous Implantation of a Ferrocene Mediated Glucose Sensor in Pigs. *Diabetologia* 29: 817.
- [3] Garguilo, M. G., Huynh, N., Proctor, A. and Michael, A. C. 1993. Amperometric Sensors for Peroxide, Choline and Acetylcholine Based on Electron Transfer Between Horseradish Peroxidase and A Redox Polymer. *Anal.Chem* 65: 523-528.

- [4] Wollenberger, U., Bogdanovskaya, V., Bordin, S., Scheller, F. and Tarasevich, M. 1999). Enzyme Electrodes using Biocatalytic Reduction of Hydrogen Peroxide. *Anal. Letters* 23: 1795-1808.
- [5] Mulchandani, A. and Wang, C. L. 1996. Bienzyme Sensors Based on Poly(anilinomethylferrocene)-Modified Electrode. *Electroanalysis* 8: 414-419.
- [6] Foulds, N. C. and Lowe. C. R. 1986. Enzyme Entrapment in Electrically Conducting Polymers. Immobilization of Glucose Oxidase in Poly-pyrrole and Its Application in Amperometric Glucose Sensors. *J. Chem. Soc., Faraday Trans. 1.* 82: 1259-1264.
- [7] Rondeau, A., Larson, N., Boujtita, M., Gorton, L. and Murr, N. E. 1999. The Synergetic Effect of Redox Mediators and Peroxidase in Bienzymatic Biosensor for Glucose Assays in FIA. *Analysis* 27: 649-656
- [8] Koide, S., and Yokoyama, K. 1999. Electrochemical Characterization of an Enzyme Electrode Based on a Ferrocene Containing Redox Polymer. *Journal of Electroanalytical Chemistry* 468: 193-201.
- [9] Calvo. E.J. and Danilowicz C. 1997. Amperometric Enzyme Electrodes. *Journal of The Brazilian Chemical Society* 6: 563-574.
- [10] Azila Abdul Aziz 2001. Amperometric Glucose Biosensors: Systematic Material Selection and Qualitative Analysis of Performance. Ph.D. Dissertation, The John Hopkins University, Baltimore, Maryland.
- [11] Liu, Y., Zhang X., Liu, H. and Deng, J. 1996. Immobilization of Glucose Oxidase onto The Blend Membrane of Poly(vinyl alcohol) and Regenerated Silk Fibroin: Morphology and Application to Glucose Biosensor. *Journal of Biotechnology* 46: 131-138.
- [12] Shuler, M. L. and Kargi, F. 2nd ed. 2002. *Bioprocess Engineering Basic Concepts*. USA: Prentice Hall.
- [13] Brooks, S. L., Ashby, R. E., Turner, A. P. F., Calder, M. R. and Clarke, D. I. 1984. Development of an On-line Glucose Sensor for Fermentation Monitor. *Biosensors* 3: 45-56.

Stability and Demulsification of Water-in-Crude Oil (w/o) Emulsions Via Microwave Heating

¹Abdurahman.H.Nour and ¹Rosli, Mohd.Yunus

¹Faculty of Chemical and Natural Resources Eng.
Universiti Teknologi Malaysia
81310 Skudai, Johor-Malaysia.

Abstract

Formation of emulsions during oil production and processing is a costly problem, both in terms of chemicals used and production losses. Experimental data are presented to show the influences of Triton X-100, Low sulphur Wax Residue (LSWR), Sorbitan monooleate (Span 83), and Sodium Dodecyl Sulphate (SDDS) on the stability and microwave demulsification of emulsions. It was found that emulsion stability was related to some parameters such as, the surfactant concentrations, water-oil phase ratio (10-90 %), temperature, and agitation speed. For economic and operational reasons, it is necessary to separate the water completely from the crude oils before transporting or refining them. In this regard, this work found that microwave radiation method can enhance the demulsification of water-in-oil (w/o) emulsions in a very short time compared to the conventional heating methods.

Keywords: Microwave, demulsification, stability, w/o emulsions, heating, surfactants.

1. Introduction

As crude oil is always produced with water, many problems occur during oil production because of the formation of emulsions (Schramm, 1992). Emulsions are difficult to treat and cause a number of operational problems such as tripping of separation equipment in gas-oil separating plants, production of off-spec crude oil, and creating high pressure drops in flow lines (Sunil, 2002). There are two forms of emulsions: water-in-oil (w/o), and oil-in-water (o/w). Most common emulsions in the oil field are water-in-crude oil (w/o) emulsions. Stability is an important characteristic of a water-in-oil emulsion. Characterization of an emulsion as stable or unstable is required before other properties can be considered, because properties change significantly for each type of emulsion.

For economic and operational reasons, it is necessary to separate the water completely from the crude oils before transporting or refining them. Minimizing water levels in the oils can reduce pipeline corrosion and maximize pipeline usage (Taylor, 1992). Effective separation of crude oil and water is an essential operation in order to ensure not only the quality of crude oil but also the quality of the separated water phase at the lowest cost (Dalmazzone, et al, 2005). The concept of microwave demulsification of emulsions was first introduced by (Klaila, 1983) and (Wolf, 1986). Authors (Fang, et al, 1995) presented demulsification of water-in-oil emulsions used microwave heating and separation method.

Corresponding Author Email: nour2000_99@yahoo.com

The objective of this paper is conducted to examine the influences of triton-X-100, Span 83, LSWR, and SDDS on emulsion stabilization and microwave demulsification of water-in-crude oil emulsions. Our findings showed that emulsion stability is related to surfactant concentrations, stirring time, temperature, water-oil phase ratio, and agitation speed. The demulsification rate was significantly accelerated by microwave radiation.

2. Materials and Methods

Two types of crude oil were used: Crude Oil A from Iran oilfield and Crude Oil B from Malaysia oilfield. Their respective compositions are given in **Table 1**. For emulsion preparations, a distilled water was used as the water phase (dispersed phase), and crude oil as oil phase (continuous phase) for both crude oil A and B. The commercially available Triton X-100, low sulphur wax residue (LSWR), sodium dodecyl-sulphate (SDDS), and sorbitan monooleate (Span 83) were used to emulsify the emulsion in this work. Emulsions were prepared in 900 ml graduated beaker, with ranges by volume of the water and oil phase. The emulsions were agitated vigorously using a standard three blade propeller at speed of 1600 rpm and temperature of 28 °C for 7 minutes. The concentrations of water in samples were 10-90 % by volume. The composition of W/O emulsion formulations and their corresponding stabilities are given in **Table 2** which shows the surfactants used for this study. In order to prepare water-in-oil (w/o) emulsions, the agent-in-oil method was followed; that is in this work, the emulsifying agents were dissolved in the continuous phase (oil), then water was added gradually to the mixture. The volume of water settled to the bottom was read from the scale on the beaker with different times. The amount of water separation in

percent was calculated as separation efficiency (e) from volume of water observed in the beaker as follows:

$$(\% \text{ of water separation, } e) = \frac{(\text{Volume of water layer, ml})}{(\text{Original amount of water, ml})} * 100 \% \quad (1)$$

The demulsification experiments were performed using Elba domestic microwave oven model: EMO 808SS, its rated power output is 900 watts and operation frequency of 2450 MHz. A 900 ml graduated cylindrical glass was used as sample container. The diameter and height of emulsion sample container were 11.5 cm and 11 cm respectively. Three thermocouples type (K-IEC-584-3) were connected to Pico-TC-08 data logging, and connected to microwave oven as shown in figure 1. The data logging was connected to PC; with PicoLog R5.08.3 software. The thermocouples were inserted to different locations, top, middle, and bottom of the emulsion sample to measure local temperatures.

Table1- Density, viscosity, surface tension, and interfacial tension of crude oils.

Density (g/cm ³)	Crude Oil A	Crude Oil B
		0.852
Viscosity (cp)	20.75	9.42
Surface tension (mN/m) at 28 °C	28.20	26.20
Interfacial tension (mN/m) at 28 °C	27.30	24.00

Table-2 Composition of W/O emulsion formulations and their corresponding stabilities

Emulsion	Stabilizer	Wt % Stabilizer In Ext. Phase (oil)	Emulsion Type	% Internal Phase
1	LSWR	0.50	w/o	50
2		0.75	w/o	55
3		2.00	w/o	60
4		5.00	w/o	70
5	Triton X-100	0.75	w/o	50
6		0.90	w/o	55
7		1.50	w/o	60
8		3.00	w/o	70
9	SDDS	1.00	w/o	65
10		1.50	w/o	70
11		3.00	w/o	75
12		3.50	w/o	80
13	Span 83	1.50	w/o	50
14		4.50	w/o	55

Emulsion	Stabilizer	Wt % Stabilizer In Ext. Phase (oil)	Emulsion Type	% Internal Phase
15		6.00	w/o	60
16		6.80	w/o	70

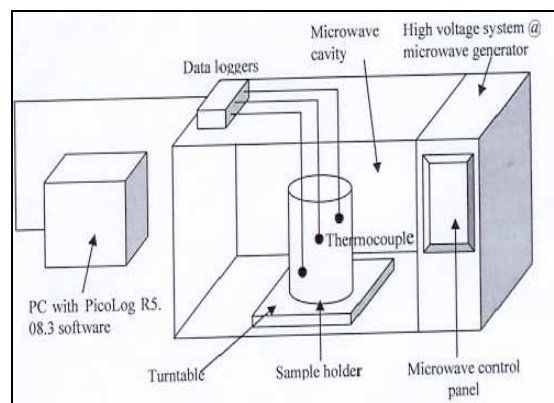


Figure1-Elba microwave oven

Table 3 Experimental results of Microwave Radiation (Crude Oils A and B)

Time (sec)	Crude Oil A		Crude Oil B	
	Temp. Increase $\Delta T, T_0=26\text{ }^\circ\text{C}$	Rate of Temp. Increase (C/s)	Temp. Increase $\Delta T, T_0=26\text{ }^\circ\text{C}$	Rate of Temp. Increase (C/s)
20	4.60	0.230	5.40	0.270
40	8.00	0.200	9.60	0.240
60	11.34	0.189	12.60	0.210
80	14.64	0.183	15.04	0.188
100	17.40	0.174	17.90	0.179
120	20.40	0.170	20.88	0.174
140	23.52	0.168	23.80	0.170
160	25.99	0.162	26.72	0.167
180	28.44	0.158	28.98	0.161
200	30.80	0.154	31.40	0.157
220	33.00	0.150	33.66	0.153
240	35.04	0.146	36.00	0.150
260	36.92	0.142	38.48	0.148

Microwave Radiation

A number of studies were carried out on microwave heating (MW) of oil and water systems. Microwave heating because of its volumetric heating effects, offers a faster processing rate. The separation of emulsified water from crude oil has several stages, due to gravity settling, water droplet/droplet flocculation takes place as water droplets approach each other (Young et al., 1996). The purpose of heating water-in-oil emulsions with microwave radiation is to separate water from oil. When water-in-oil emulsion is heated with microwave radiation, two phenomena will occur; the first one is the increase of

temperature, which causes reduction of viscosity and coalescence. The result is separation of water without addition of chemicals (Fang et al., 1988, 1989). According to Stoke's law, if oil is the continuous phase, the settling velocity of water droplets is given by:

$$v_w = \frac{(\rho_w - \rho_o)gD^2}{18\mu_o} \quad (2)$$

where D is the diameter of the droplets. The viscosity of oil very sensitive to temperature, as temperature increases, viscosity decreases much faster than the density difference, $(\rho_w - \rho_o)$ does, the result when viscosity decreases, the droplets size increases. Therefore, microwave heating increases the velocity of water (v_w), and accelerates the separation of emulsion. The second phenomenon is coagulation. The higher temperature and lower viscosity make the coagulation process easier. The results are larger particle diameter D, and rapid separation.

3. Results and discussions

The first part of our paper studied the formation, production, and stabilization of w/o emulsions, while in the second part discusses the emulsion breaking (demulsification) of w/o. According to Table 2, LSWR and Triton X-100 water-in-oil (w/o) emulsions were made with 50, 55, 60, and 70 % (v/v) internal phase. For the 50 and 55 emulsion, a higher solids concentration was found (6 mg/ml oil versus 2 mg/ml) allowed easier emulsification and slowed the settling process. The 60 % and 70 % emulsions were appeared fairly stable with little settling. The LSWR and Triton X-100 stabilized emulsions exhibited only slight coalescence over three days. Some globule formation was observed and settling occurred. In contrast, the Span 83 emulsions were different from LSWR and Triton X-100 emulsions, even at similar dispersed phase volume fractions. Emulsions were made at 50 and 55 % (v/v) internal phase with surfactant concentrations in the oil phase of 1.5 % and 4.5 % (w/w), at 60 % (v/v) (3 % w/w), and at 70 % (v/v) (6.8 % w/w). The different between surfactant concentrations for the 50 % and 55 % emulsions made by Span 83 appeared very significant on emulsions stability. High Span 83 concentrations increased emulsion stability; therefore, for high concentration of Span 83, the viscosity of w/o emulsion increased considerably and the emulsion droplets lost their shape.

The effect of the disperse phase on the stability of emulsion systems also examined with sodium dodecyl sulphate, (SDDS) as the emulsifying agent. In this regards, the SDDS emulsions were made with 65, 70, 75, and 80 % (v/v) (1 to 3.5 w/w in oil) internal phase. An increase in the concentration of SDDS in oil did not cause an increase but decrease of the demulsification rate. The decrease would be induced by the increase of the surface potential of water droplets arising from the increase of the density of SDDS anion on the surface, which overcame the decrease of the surface potential by the increase of ionic concentration in oil.

The emulsion stability for crude oils A and B were examined as function of processing time and emulsifier applied. Stability evaluated via the ratio of the total water separated. The evaluation was carried out with agitation speed at 1800 rpm at 30 minutes at 26.5 °C. As demonstrated in figures 2 and 3 in most cases stability of emulsion increases with processing time.

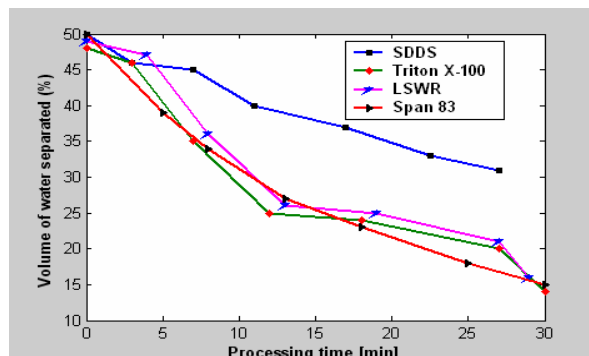


Figure 2-Change of emulsion stability for Crude oil A emulsions (50-50 % w/o), as function of processing time and emulsifier applied.

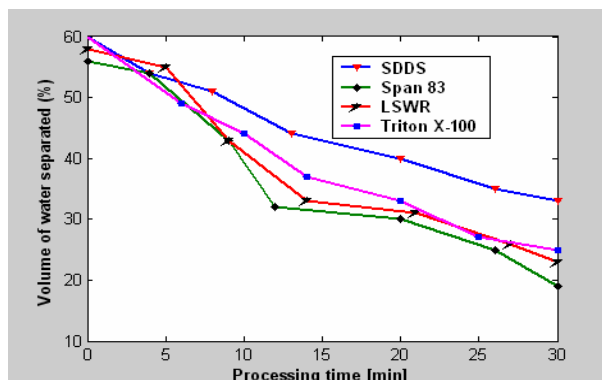


Figure 3-Change of emulsion stability for Crude oil B emulsions (50-50 % w/o), as function of processing time and emulsifier applied.

It is worth noticing that all surfactants permit a very long time for separation of the water phase (emulsion more stable). However, the maximum amount of water separated from crude oil A was (50 %) and Crude B was (60 %). From these observations, the classification in terms of decreasing stability efficiency is therefore the following; SDDS > Triton X-100 > LSWR > Span 83 respectively.

The effect of stirring time and phase ratio on emulsion stability was investigated again. A water-in-oil emulsion prepared with various volume ratios as shown in Figure 4. At low phase ratio water/oil (10/90 %) only low stability was obtained, the increment of the volume continued till (75/25 %). It's interesting to observe that, increasing the phase ratio, surfactant availability increases accordingly leading to highly stable emulsion (75/10 %). The variation in stability of the emulsions with phase ratio 75/10 % it is very difficult to explain its behavior especially during the first 8 minutes of processing. When the volume of dispersed phase reached to (90/10 %), the emulsion

behavior completely has changed Figure 4, the emulsion changed from w/o to o/w. from these measurements and observations, it can be deduced that the phase inversion point should be in the range of 68-72 % water. The oil-in-water emulsion with a phase ratio of 90/10 % is very instable emulsion. As the volume of the dispersed increases, the continuous phase must spread out farther to cover all of the droplets. This causes the likelihood of impacts to increase, thus decreasing the stability of the emulsion. This means that, the emulsion might not break as increase the volume of the dispersed phase. In fact this increment caused an emulsion to invert from one phase (w/o) to another (o/w).

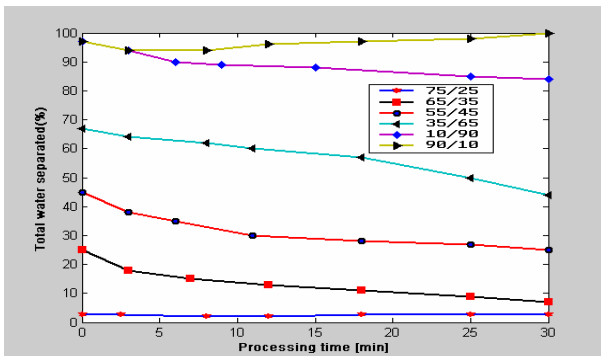


Figure 4-Change of (w/o) emulsion stability for crude oil emulsions (stabilized with Span 83 surfactant) as function of processing time. Stability evaluated via the ratio of total water separated.

The above section was discussed the formation, production, and stabilization of emulsions. This section, discusses the vise versa of emulsion stability which is breaking emulsions (demulsification). Table 3 shows the experimental results of microwave radiation for crude oil emulsions A and B. it is obvious that, there were a correlation between the radiation time and the rate of temperature increase, as radiation time increases, the rate of temperature (dT/dt) decreases figure 5. Also the rate of temperature increase (dT/dt) decreases at higher temperatures, this may attributed due to the small dielectric loss of water. The average rates of temperature increase for crude oil emulsion A and B were found as 0.171 and 0.182 C/sec respectively, the same findings were reported by (Fang et al, 1995). Since the purpose of heating water-in-oil emulsions with microwave is to separate water from oil, therefore, the separation efficiency of Crude oil A and B emulsions calculated by using Equation (1) were shown in Figures 6 and 7 respectively.

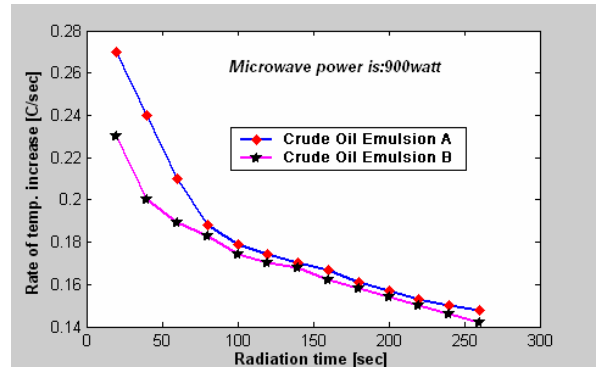


Figure 5 Heating rate vs. Microwave radiation time for Crude Oil Emulsion (A & B)

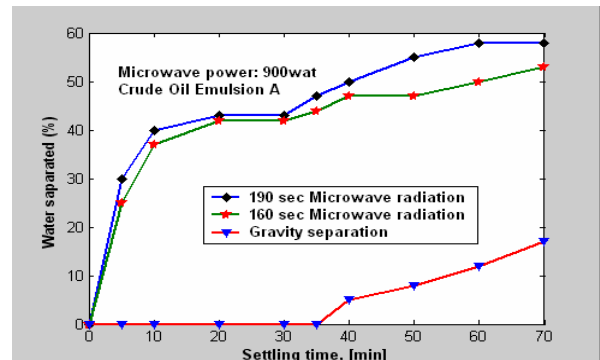


Figure 6-Separation of water from Crude Oil Emulsion A.

As shown in figure 6; the separation in much faster with microwave heating than room temperature (gravity separation). Figure 6 shows that at the end of 35 minutes of gravity settling, there is no separation of water layer was observed. For crude oil A, at microwave radiation times 160 and 190 seconds, the separation of water was found 54 % and 59 %, respectively. While for crude oil B, as shown in figure 7 at radiation times 140, 160, and 190 seconds, water separation were found 50 %, 57 % and 68 % respectively.

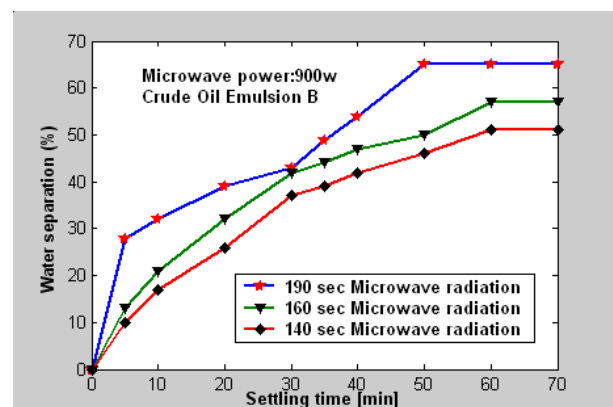


Figure 7-Separation of water from Crude Oil Emulsion B.

Results of table 3, with combination of figures 6 and 7 showed microwave radiation can raise the temperature of emulsion, reduce viscosity, and the result a separation of water from oil as mentioned by Equation (2).

4. Conclusions

The water-in-crude oil emulsion has great importance in the oil industry. The formation, production and stabilization of water-in-crude oil (w/o) emulsions are investigated over a wide range of parameters. These parameters are surfactant concentrations, temperature, stirring time, water-oil phase ratio (10-90 %), and agitation speed (800-1800 rpm). In terms emulsion breaking, the microwave demulsification was applied on water-in-oil emulsions. Results showed that microwave radiation is a dielectric heating technique with the unique characteristics of fast and volumetric heating. The microwave separation does not require chemical addition.

References

1. Dalmazzone, C., Noik, C., and Komunjer, L. (2005). Mechanism of Crude Oil-Water interface Destabilization by Silicone Demulsifiers. March 2005. Journal of SPE. Pp.44-49.
2. Fang, C.S., Chang, B.K.L., and Lai, P.M.C. (1988). Microwave Demulsification. Chem. Eng. Comm. Vol.73.pp.227-239.
3. Fang, C.S., Lai, P.M.C., and Chang, B.K.L. (1989). Oil Recovery and Waste Reduction by Microwave Radiation. Environmental Progress. Vol,8,no.4.pp.235-236.
4. Fang, C.S., and Lai, P.M.C. (1995). Microwave Heating and Separation of Water-in-Oil Emulsions. Journal of Microwave Power and Electromagnetic Energy. Vol.30.No.1.pp.46-57.
5. Klaila, W.J. (1983). Method and apparatus for Controlling Fluency of high viscosity hydrocarbon fluids. U.S. Patent. 4,582.629.
6. Schramm, L.L. (1992). Emulsions: Fundamentals and Applications in the Petroleum Industry. Adv.Chem.Ser. 231. ACS. Washington, DC.
7. Sunit, K. (2002). Crude Oil Emulsions: A State of-The-Art Review. Society of Petroleum Engineers Inc. San Antonio, Texas.
8. Taylor, S.E. (1992). Emulsions: Fundamentals and Applications in Petroleum Industry. Chem.Ind.Vol.20, pp. 770-773.
9. Wolf, N.O. (1986). Use of Microwave Radiation in Separating Emulsions and Dispersions of Hydrocarbons and water. U.S. Patent. 4,582.629.
10. Young, H.K., Nikolov, A.D., Wasan, D.T., Diaz-Arauzo, H., and Shetty, C.S. (1996). Demulsification of Water-in-Crude Oil Emulsions: Effects of film Tension, Elasticity, Diffusivity, and Interfacial Activity of Demulsifier Individual Components and Their Blends. Dispersion Science and Technology, 17(11).Pp 33-53.

Supercritical Fluid Carbon Dioxide for the Extraction of Polychlorinated Biphenyls in Sewage Sludge

Umi K. Ahmad and Ahmad Zamani Ab Halim

Department of Chemistry, Faculty of Science, Universiti Teknologi Malaysia
81310 UTM Skudai, Johor, Malaysia.
Tel: 07-5534522, e-mail: umi@kimia.fs.utm.my

Abstract

Extractions of organic contaminants such as polychlorinated biphenyls (PCBs) from sewage sludge have often involved the use of large volume of organic solvents and long extraction time. In this study, a supercritical fluid CO₂ (SF-CO₂) was utilised as the extractant developed with the aim of overcoming the disadvantages associated with other conventional extraction technique such as soxhlet extraction. The optimum extraction pressure, temperature, time, flow rate, cosolvent composition and solvent trapping system that enable maximum extraction of seven PCBs (PCB 28, 52, 101, 118, 138, 153 and 180) were determined based on analyte recoveries and extracts were analyzed by gas chromatography–electron capture detector on a HP-5MS capillary column. Recoveries of 60-99 % were obtained for the seven PCB congeners analysed. Comparison of PCBs extracted with Soxhlet extraction revealed the SFE technique to be fast, easy, and environmentally clean technique.

Keywords: Polychlorinated Biphenyls, Sewage Sludge, SF-CO₂, Soxhlet, GC-ECD

1. Introduction

Disposal of municipal sewage sludge and wastes have become a major problem in many countries in the world due to rapid population growth and urbanization. Malaysia is faced with the same problem. Approximately 3 million cubic meters of sewage sludge is produced annually in Malaysia and the total cost of managing it was estimated at RM one billion (Kadir and Haniffa, 1998). By the year 2020, the volume is estimated to increase to 7 million cubic meters, which will require about double the KLCC twin tower to fill or almost 1.4 million tanker trips to manage (Velayutham and Kadir, 1999).

Sewage sludge contains many organic contaminants that may have an impact on soil and the food chains associated with arable land. Polychlorinated biphenyls (PCBs) have been produced worldwide or used as dielectric fluids in capacitors and transformers, hydraulic fluids, fire retardants, and plasticizers (Hutzinger *et al.*, 1974). For several of these applications, PCBs are chosen because of their physical and chemical stability and their electrical insulating properties. Due to their weak solubility, PCBs are eliminated from water in the process of partitioning during sewage treatment.

Traditionally, PCBs determination has been performed by Soxhlet extraction technique. This technique is often

time-consuming and requires large volumes of organic solvent. The use of large volumes of extracting solvents adds additional cost because of the fee associated with purchasing and disposal of toxic solvents. For these reasons, alternative extraction strategies have been developed and applied namely supercritical fluid CO₂ extraction.

SFE has attracted intense interest during the past 20 years, mainly for extraction of solid samples, because it offers short extraction times and minimum use of organic solvents. A supercritical fluid is defined as substance above its critical pressure and temperature, resulting in a fluid that is more penetrative and of higher solvating power (Castro. *et al.* 1994). CO₂ has been popular because of its low cost, availability and safety, and its suitable critical temperature (31.2 °C) and pressure (72.8 atm). The optimum pressure, temperature, time factors, flow rate, cosolvent composition and trapping system that enable maximum extraction of seven PCBs (PCB 28, 52, 101, 118, 138, 153 and 180) from sewage sludge were determined in this study.

2. Experimental

2.1 Reagents

Neat individual standards of PCB 28, 52, 101, 118, 138, 153 and 180 (Figure 1) at a concentration of 10 µg/mL were obtained from Dr. Ehrenstorfer Lab (Germany).

* Corresponding Author. E-mail: umi@kimia.fs.utm.my,
Tel: +60-7-5534522

A mixed standard solution was prepared in *iso*-octane (BDH Limited Poole England) at a final concentration of 1.0 µg/mL, which was used to spike the sludge and for calibration purposes Methanol was obtained from Merck (Germany), *iso*-octane from BDH Limited (Poole, England), whereas *n*-hexane was obtained from Fisher Scientific (USA). All other metal salts and reagents were analytical grade. Liquid CO₂ was industrial grade from NISSAN IOI Berhad (Pasir Gudang, Johor)

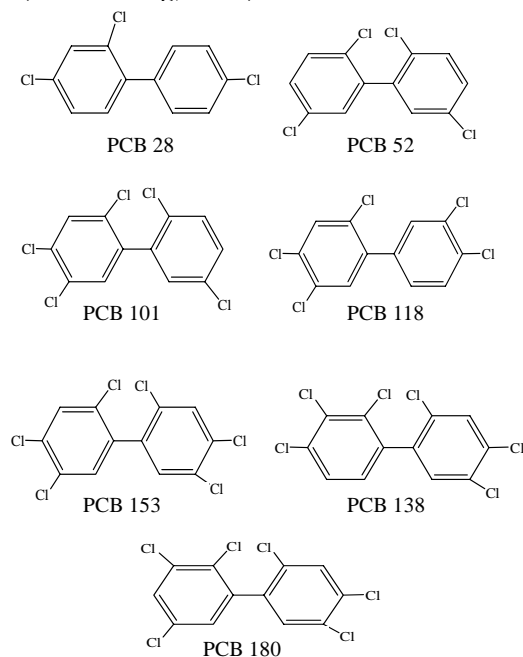


Figure 1 Molecular Structures of seven PCBs under study

2.2 Instrumentation and Apparatus

The schematic set-up of the extraction unit is shown in Figure 2. The SFE apparatus consisted of a Jasco PU-1580 intelligent HPLC pump (Jasco Cooperation, Tokyo Japan) fitted with a cooling jacket to deliver CO₂. In order to cool the pump head, ethylene glycol and water mixture (50:50, v/v) was circulated through the cooling jacket using a chiller and circulator, which delivered the coolant down to -6 °C. A 5 mL extraction cell was used as the extraction vessel. A Jasco CO-1560 Intelligent Column Oven (Jasco Cooperation, Tokyo Japan) was used to maintain a constant temperature of 60 °C while a dynamic mixer (Jasco HG1580-32) was utilised for mixing organic modifier and liquid CO₂. A Jasco back pressure regulator (BP-1580-81) was used to control the extraction pressure.

PCBs analyses were carried out using a Perkin Elmer XL GC equipped with an electron capture detector (ECD) and data processing using Chromatography Version 4.1 software. Nitrogen was used as the make up gas for ECD at a flow rate of 30 mL/min and helium as the carrier gas at pressure of mL/min. An HP-5MS column (30 m x 0.25 mm ID x 0.25 µm film thickness) from Hewlett Packard, (USA) was used. Injection port temperature was set at 280 °C and detector temperature at 290 °C. The temperature was initially held at 200 °C for 3 minutes, and then programmed from

200 °C to 230 °C at a ramp of 5 °C/min. The temperature was again held for 3 minutes at 230 °C and then increased to a further 260 °C at 2 °C/min.

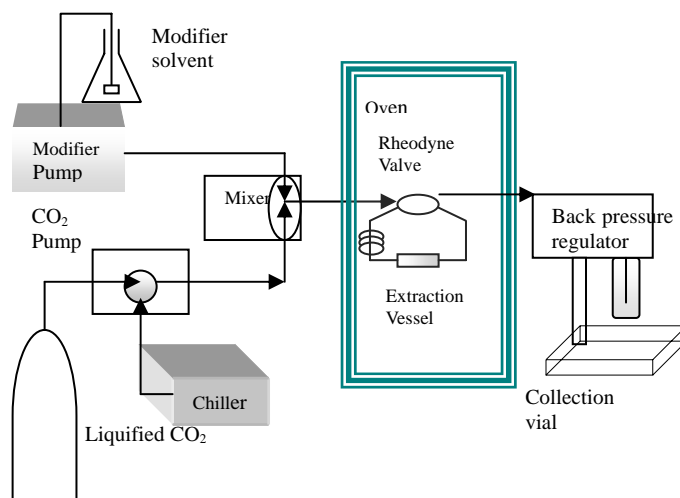


Figure 2 Schematic diagram of the basic set up for SFE

2.3 Sewage Sludge Samples

Blank sludge samples used for analyte recovery studies were prepared in the laboratory. The sludge samples were initially subjected to exhaustive Soxhlet extraction for removal of organic contaminants, therefore enabling it to be used as blank sludge samples. The blank sludge sample was accurately weighed and spiked with PCB standard mixture. The spiked sample was left to expose at room temperature in order to evaporate off the solvent.

2.4 Supercritical Fluid Extraction

The blank sludge sample was filled into the extraction vessel (5.0 mL) and spiked with 100 µL of the PCB mixture (1.0 mg/L). The spiked sludge sample was exposed at room temperature in order to evaporate off the solvent. Prior to extraction, the SFE oven system was preheated to equilibrate the system. The sludge was then extracted with supercritical carbon dioxide using SFE/SFC Jasco instrumentation. Pure carbon dioxide was selected for SFE for preliminary extractions of the blank sludge. Analyte collection was performed through a stainless steel outlet in a solvent trap placed in the collection vial. The extract was then evaporated to near dryness and 100 µL *iso*-octane was added to the extract prior to injection into the GC-ECD.

2.5 Conventional Soxhlet Extraction

Spiked sludge sample (1.0 g) was placed in an extraction thimble together with 1.0 g of anhydrous sodium sulfate. Soxhlet extraction was performed using 320 mL of acetone-*n*-hexane (1:1) mixture and the extraction carried out for 24 hours. The soxhlet extract was filtered through a Whatman filter paper. The filtrate was concentrated to about 1-2 mL using a rotary evaporator. The elimination of sulfur

was carried out using copper powder (0.5 g) previously activated using concentrated HCl. The final determination by GC-ECD was carried out.

3. Results and Discussions

3.1 Gas Chromatographic Separation of PCB mixture

Seven PCBs of common occurrence in sewage sludge were chosen for determination with GC-ECD using a HP-5MS capillary column. All seven PCBs were successfully separated, with a total time of less than 30 minutes. The gas chromatographic resolution of seven PCBs is illustrated in Figure 3.

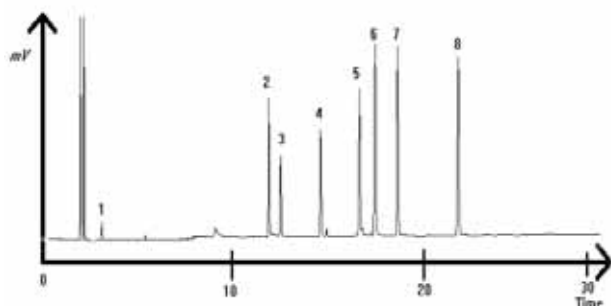


Figure 3 GC-ECD separations of the PCBs using a HP-5MS capillary column. Peak identity: (1) *iso*-octane (2) PCB 28, (3) PCB 52, (4) PCB 101, (5) PCB 118, (6) PCB 153, (7) PCB 138, and (8) PCB 180

Table 1 shows the capacity factors of the PCBs. The PCBs were found to elute according to the increase in their molecular weight and chlorine atom position as well as their boiling points.

Table 1 Capacity factor of the standard PCBs

PCB	Molecular Formula	t_r (min)	k'
PCB 28	2, 4, 4' - triCB	11.71	1.19
PCB 52	2,2', 5, 5' - tetraCB	12.33	1.44
PCB 101	2, 2', 4, 4', 5, 5' - pentaCB	12.71	2.22
PCB 118	2, 3', 4, 4', 5 - pentaCB	14.32	3.16
PCB 153	2, 2', 4, 4', 5, 5' - hexaCB	17.17	3.51
PCB 138	2, 2', 3, 4, 4', 5' - hexaCB	18.32	3.93
PCB 180	2, 2', 3, 4, 4', 5, 5' - heptaCB	21.41	5.30

* $k' = (t_r - t_0) / t_0$, CB = ChloroBiphenyl, t_r = Retention time, t_0 =

3.2 SFE Recoveries of PCBs

SFE was initially considered as the new extraction technique of choice for the extraction of PCBs for sludge. In order to evaluate the efficiency of the supercritical CO₂ extraction of PCBs, recovery studies of PCBs were conducted using spiked sludge samples. Certified Reference Material for sludge was difficult to obtain, so these studies were conducted using spiked blank sludge.

To obtain the optimum SFE conditions for the extraction of PCBs in sludge sample, several parameters

such as extraction temperature, extraction pressure, composition of extractant modifier, CO₂ flow-rate, solvent trapping and extraction duration time were optimized.

3.2.1 Effect of Extraction Temperature

The effect of extraction temperature was initially observed by varying the extraction temperature (40 to 70 °C) while keeping the other parameter constants (extraction time of 30 minute, CO₂ flow rate of 2.0 mL/min, extraction pressure of 20 Mbar and analyte collected in methanol as the trap solvent). The recoveries of seven PCBs mixture (spiked at 1.0 mg/L) were investigated.

As shown in Figure 4, all the PCB recoveries generally increased as the temperature increased. At an extraction temperature of 60 °C, all PCBs gave over 80 % recoveries with the exception of PCB 28 and PCB 52. On further increment to 70 °C, recoveries of most PCBs were not significantly different. Previous work indicated that pure CO₂ at high temperature could yield high recoveries of many semi-volatile pollutants (Langenfeld *et al.*, 1993; Yang *et al.*, 1995; Kreuzig *et al.*, 2000 and Sporryng *et al.*, 2005). However in this study, an extraction temperature above 70 °C was not utilised due to the maximum upper temperature of the oven being at 80 °C.

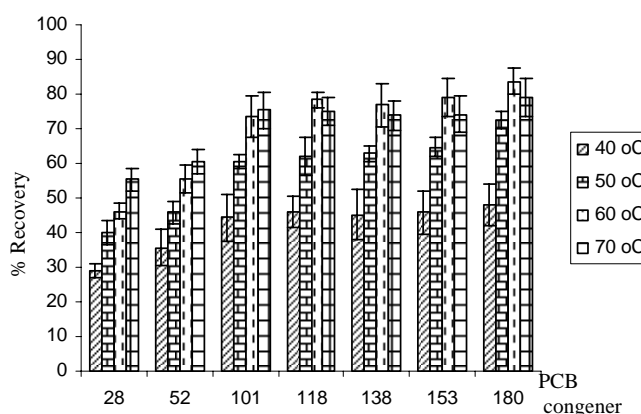


Figure 4 Effect of extraction temperature on the SFE recovery of the spike sludge using pure CO₂ at 2.0 mL/min with 30 min extraction and extraction pressure of 20 Mbar.

An increased recoveries at high temperature extractions could be contributed by two factors (Castro *et al.*, 1994; Taylor,1996). Firstly, high temperature can produce higher vapor pressure of the analytes, thus increasing the analytes solubility despite a decrease in the density of CO₂ at high temperature. Secondly, a certain amount of energy is required to desorb the analyte from sludge. Hence, an increase of the extraction temperature can provide more energy to the extraction system. This will also increase the rate of the desorption process. The temperature may also affect the diffusion as well as the mass transfer behavior. Extractions at high temperature may change the pores and increase contact area of sludge with CO₂.

The reproducibility of extraction temperature on SFE was found to be comparatively good for all samples with RSD below 10 %. Based on the above results, the optimum

temperature was taken as 60 °C.

3.2.2 Effect of Extraction Pressure

Using the blank sludge as a model sample matrix, the PCB recoveries were studied over the pressure ranging from 10 to 40 Mbar by 10 Mbar increments, at a constant extraction temperature of 60 °C within 30 min extraction. The maximum operating pressure of the back pressure regulator for the SFE system used was 50 Mbar. The effect of pressure on the extraction efficiency of PCBs is shown in Figure 5. The best optimum pressure for the extraction of PCB from sludge was found to be 20 Mbar which yielded the recoveries of PCB 101 to PCB 180 in the range of 70 to 81 %. However, PCB 28 and PCB 52 only yielded recoveries of 45 and 55 % respectively.

Generally when extraction pressure increases from 10 to 40 Mbar, the density of CO₂ also increases. The extraction pressures up to 30 Mbar caused a large decrease in the extraction efficiency for all PCBs except for PCB 28 and PCB 52. Further increase of pressure (40 Mbar) gave reduced recoveries of all PCBs.

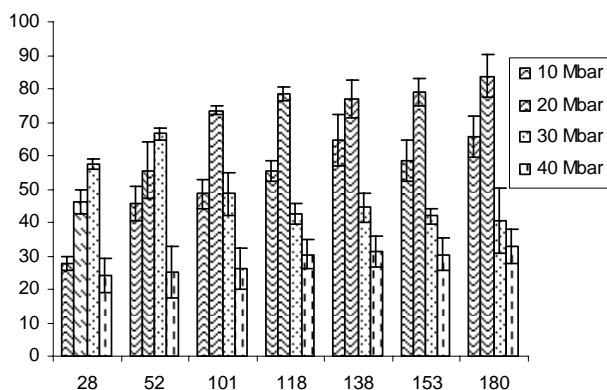


Figure 5 Effect of extraction pressure on the SFE recovery of the spiked sludge sample using CO₂ at 60°C with 30 min extraction and CO₂ flow rate of 2.0 mL/min.

Thermodynamic properties such as solute solubility are necessary to understand the pressure effect on the extraction process (Langenfeld *et al.*, 1993). Velde *et al.*, (1992) reported that, the largest improvement in recovery was seen when increasing the pressure from 15 to 20 Mbar, whereas only a slight increase in recovery is obtained from 25 to 20 Mbar. The solvating power of SF-CO₂ increases proportionally with density, but the diffusion coefficients are inversely proportional to higher densities. This result obtained in this study is similar to that reported by Tong and Imagawa (1995) where the best optimal pressure for extraction of PCBs was found to be 20 Mbar.

At high pressures, the risk of coextraction increase from the matrix is possible, so a moderate extraction pressure at 20 Mbar was adequate for extraction of PCBs in sludge. The reproducibility of extraction pressure on SFE was shown to be comparatively good for all samples with RSD below 9.6 %.

3.2.3 Methanol as Organic Modifier

Previous studies have indicated that the solvating power of pure SF-CO₂ is not strong enough to extract persistent pollutant like PCBs from environmental samples and the use of cosolvent is necessary to enhance the recoveries in SFE (Chen *et al.*, 1997; Abaroudi *et al.*, 1999). A few percent (1 to 15 %) of a polar modifier is often added to enhance the solubility of more polar compounds or to increase the ability of the supercritical fluid to better displace analytes from the matrix active sites. These modifiers are usually organic solvents that are added to the extraction fluid. They can be added using three different procedures (Castro *et al.*, 1994; Taylor, 1996). The modifier may be directly added to the sample. Premixed fluids may be prepared by adding a modifier to the liquid CO₂ in the storage cylinder or by using a separate modifier pump to pump in the organic solvent into the SF-CO₂ fluid.

Figure 6 shows the chromatogram of PCBs separation with and without modifier during the extraction. The separation using 15 % methanol as modifier recorded more peaks on the chromatogram compared to the one without modifier. The color of the SFE extract changed to pale yellow using methanol modified CO₂ which indicated that some components of sludge were extracted during SFE.

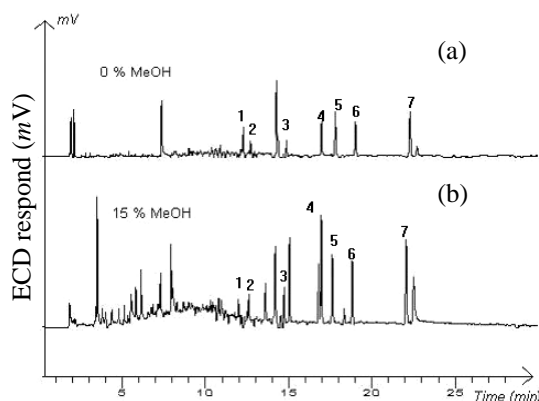


Figure 6 GC-ECD separation of PCBs (a) without modifier (neat SF CO₂) and (b) SF CO₂ with 15 % methanol as modifier. Peak identity: (1) PCB 28, (2) PCB 52, (3) PCB 101, (4) PCB 118, (5) PCB 153, (6) PCB 138, and (7) PCB 180

From Figure 7, when the compositions of methanol increased from 0 to 10 %, the recoveries were found to increase slightly were for all PCBs. SF-CO₂ with 15 % methanol extraction gave the best overall extraction efficiency for all PCBs except for PCB 28 with only 49 % recovery. The increase composition of methanol to CO₂ does not give any significant changes for PCB 28. The percent recoveries of PCB 28 are lower than any other PCB due to lower formula molecular weight or lost during trapping system. Based on the results, the optimal percent of adding methanol to SF CO₂ was taken at 15 % methanol. The reproducibility through the application of extraction modifier on SFE was shown to be comparatively good with all samples with RSD below 10.2 %.

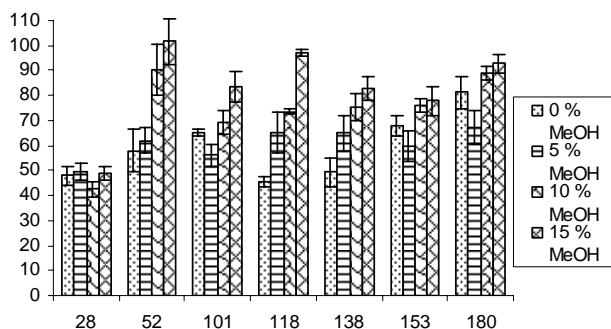


Figure 7 Effect of varying concentration of methanol as modifier on the SFE recoveries PCBs. SFE conditions: pressure 20 Mbar, SF flow rate 2.0 mL/min, 60°C with 30 minute extraction time

3.2.4 Flow Rate

The effect of varying flow rate on the extraction efficiency of PCBs in spiked sludge was studied. The composition of flow rate was varied from 1.0 mL/min to 2.5 mL/min. From Figure 8, when SF-CO₂ flow rate was increased from 1.0 mL/min to 1.5 mL/min, recovery of all PCB was found to be insignificant. When SF CO₂ was increased to 2.0 mL/min, the recovery of all PCB increases, however on further increment to 2.5 mL/min, the recovery all PCB decreases. High flow rate could decrease the recovery either by inducing an elevated pressure drop through the extraction cell or by increasing analyte loss during decompression of the supercritical fluid.

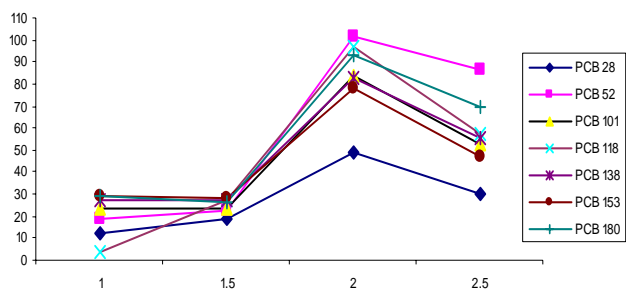


Figure 8 Effect of varying flow rate on the SFE recoveries PCBs. SFE conditions: pressure 20 Mbar, SF CO₂ with 15 % methanol as modifier and 60 °C with 30 minute extraction time.

However, only PCB 28 showed the lowest recovery (50 %) at the optimum flow rate 2.0 mL/min due to PCB 28 be having polarity lower compared to other isomer. Based on the results in Figure 8, it was shown that the flow rate of 2.0 mL/min gave the best overall extraction efficiency for all PCBs. The data reproducibility using extraction flow rate on SFE was shown to be comparatively good for all samples with RSD below 9.2 %.

3.2.5 Effect of Trapping Solvent

As the analyte solubility depends on the SFE trapping solvent, three solvents of different polarity and boiling points, namely isooctane, methanol and n-hexane were examined for their efficiency as trapping solvents for PCBs extraction. It was found that trapping solvents showed varying efficiency as shown in Figure 9. Generally, methanol and n-hexane were found to give better analyte recovery than iso-octane, although iso-octane showed higher recovery for PCB 28 and 118 compared to methanol and n-hexane. The reproducibility of extraction solvent trapping on SFE was shown to be comparatively good with all samples displaying RSD below 9.2 %.

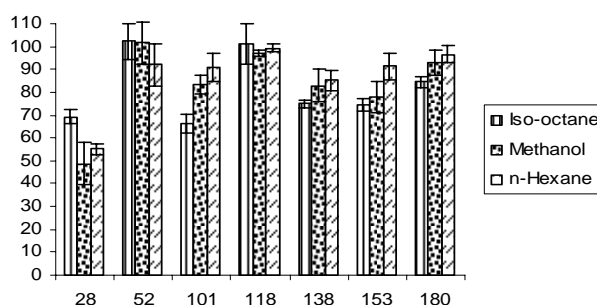


Figure 9 Effect of different solvent trapping on the SFE recoveries of PCB congener. SFE conditions: pressure 20 Mbar, temperature 60 °C, 15 % methanol as modifier and with 30 minute extraction time.

3.2.6 Duration of Extraction Time

Figure 10 shows the extraction profile of online SFE-UV for PCBs at the optimum temperature 60 °C and pressure of 20 Mbar using 15 % methanol as modifier. From the extraction profile, it can be seen that the profile increased sharply as the extraction began, and later decreased indicating that the PCBs were extracted from the matrix. A 15 minute extraction time was found sufficient to extract the PCBs from the spiked sample. As for the extraction of real samples, a longer time is needed to force the analytes out from the matrix site and the time of extraction was doubled to 30 minutes.

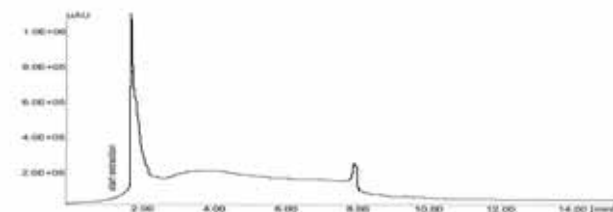


Figure 10 Extraction profile for PCBs using a temperature of 60 °C, pressure of 20 Mbar and 15 % methanol as modifier.

3.2.7 Comparison of Extraction Technique

Table 2 listed the comparison of the techniques applied for the extraction of PCBs. The organic Soxhlet extracts dark

yellow and highly turbid, whereas the SFE extracts (collected in *n*-hexane) were pale yellow with little turbid. The organic soxhlet extracts also yielded more artifact peaks in the GC-ECD chromatograms compared to the SFE extract.

Table 2 Comparison of extraction technique

	Soxhlet	SFE
Extraction Solvent	Dichloromethane-Acetone	SF-CO ₂ + Methanol modifier
Collection Solvent	-	<i>n</i> -hexane
Volume of organic solvent use	Large (320 mL)	Small (10 mL)
Pressure	Ambient	20 Mbar
Temperature (°C)	Boiling point of solvent	60
Extraction time	24 hour	30 minute
Color extract	Dark yellow	Pale yellow
Extract clean-up	Required	Not required

Large volumes of solvent are necessary for soxhlet extraction compared to SFE. The need to dispose of these solvents presents the potential for further environmental contamination and added expense. Soxhlet extraction is also very time consuming compared to SFE method that only needed 30 minute to extract analyte for the sample.

The PCB recovery SFE result was based on soxhlet extraction for 24 hour using *n*-hexane/acetone (1:1, v/v) is shown in Table 3, *n*-hexane is rather non polar whereas dichloromethane is a halogenated solvent. In this study, *n*-hexane/acetone was applied as extraction solvent based on previously research. From the Table 3, SFE showed good PCB recoveries from 50 to 99 % compared to soxhlet extraction. SFE also showed the added advantages of low organic solvent usage, short extraction and environmentally clean technique compared soxhlet extraction.

Table 3 Comparison of PCBs recovery using Soxhlet and SFE

PCB No	Extraction	Percent Recovery ± RSD	
		Soxhlet	SFE
28		52.2 ± 4.2	55.2 ± 2.5
52		58.2 ± 3.2	91.9 ± 9.2
101		74.1 ± 5.0	91.0 ± 6.2
118		70.8 ± 2.2	99.2 ± 1.6
138		72.3 ± 3.2	85.1 ± 4.6
153		74.4 ± 2.6	91.4 ± 5.8
180		75.2 ± 4.2	96.3 ± 3.8

4. Conclusions

Supercritical CO₂ extraction has been shown to be

effective for the extraction of the target PCBs from the spiked sludge. The optimized SFE condition used extraction temperature 60 °C, extraction pressure of 20 Mbar, 15% methanol as co solvent, extraction flow rate at 2.0 mL/min, 30 minute extraction and *n*-hexane as the trapping solvent. SFE yielded cleaner extracts than the Soxhlet method. SFE was found to be a viable and a good alternative to conventional soxhlet extraction for the analysis of PCBs in sludge.

5. Acknowledgement

Thanks are due to the Ministry of Science, Technology and Innovation (MOSTI) for financial support (IRPA Vote 74085; 08-06-060027 EA 204) under the IRPA RM 8 mechanism.

6. References

- [1] Hing, L., D., Zenz, D., R., Tata, P., Kuchencrither, R., Malina, J., F., and Sawyer, Jr., B. (1998). "Municipal Sewage Sludge Management : A Reference Text on Processing, Utilization and Disposal". Volume 4. second edition. Technomic Publishing Company Inc. Lancaster, Pennsylvania, USA.
- [2] Hutzinger, O., Safe, S., and Zitko, V., (1974). "The Chemistry of PCBs". CRC Press, Inc.
- [3] Kadir, A., and Haniffa, M., (1998). "The Management of Municipal Wastewater Sludge in Malaysia". Paper Presented at IEM Talk on Sewage Sludge Management Issues, 18th August 1998. Petaling Jaya.
- [4] Castro, M., D., Valcarcel, M., and Tena, M., T., (1994). "Analytical Supercritical Fluid Extraction". Springer Lab Manual. Germany.
- [5] Sporing, S., Bøwadt, S., Svensmark, B., and Björklund, E. (2005). Comprehensive Comparison of Classic Soxhlet Extraction with Soxtec Extraction, UE, SFE, MAE and ASE for the Determination of PCBs in Soil. *J. Chromatogr. A*. **1090**. 1-9
- [6] Van Der Velde, E., G., Haan, W., D., and Liem, A., K., D., (1992). Supercritical Fluid Extraction PCBs and Pesticides from Soil. *J. Chromatogr.* **626**. 135-143.
- [7] Tong, P., and Imagawa, T. (1995). Optimization of Supercritical Fluid Extraction for Polychlorinated Biphenyls from Sediments. *Anal. Chim. Acta.* **310**. 93-100.
- [8] Cheng, P., Zhou, W., and Tavlarides, L., L. (1997). Remediation of Polychlorinated Biphenyl Contaminated Soil/Sediments by Supercritical Fluid Extraction. *Environ. Prog.* **16**. 227.

Optimization of Headspace-SPME Using Lab-Made C₈-Coated Fiber for the Forensic Analysis of Arson Accelerants in Fire Debris

Umi K. Ahmad¹, Abdul Rahim Yacob², Geetha Selvaraju³

Department of Chemistry, Faculty of Science,
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
¹Tel: +07-5534522, Fax: +07-5566162 E-mail: umi@kimia.fs.utm.my
²Tel: +07-5534505, Fax: +07-5566162 E-mail: manrahim@kimia.fs.utm.my
³Tel: +07-5534531, Fax: +07-5566162 E-mail: geetha_selvaraju@yahoo.com

Abstract

Interpretation of forensic evidence and in particular, the ability to uniquely detect accelerant is a crucial challenge in the scientific investigation of arson. Limitations in this area are posing a significant impact on police investigations and the successful prosecution of arson cases. Much interest has been shown recently in replacing conventional accelerant residues extraction technique with headspace solid phase microextraction (HS-SPME) technique. In HS-SPME, the extraction is based on the enrichment of components on an adsorbent coated fused silica fiber. A lab-made fiber prepared by sol-gel method, containing 1:1 molar ratio of octyltriethoxysilane (C₈-TEOS): methyltrimethoxysilane (MTMOS) was employed in this technique. The parameters effecting the sorption of analytes onto the fiber (HS-SPME operating conditions), such as extraction time, extraction temperature and desorption time were optimized. The developed HS-SPME method using sol-gel C₈-coated fiber showed satisfactory reproducibility (RSD < 6 %) and detection limits for accelerants (0.7-1.0 µL) under the optimum experimental conditions. Electron microscopy experiments revealed that the surface of the fiber coating was well-distributed and a porous structure was suggested for the sol-gel derived C₈ coating with an approximate thickness of (3-4) µm. The chemistry behind the coating process will also be discussed. The lab-made SPME fiber was shown to be a good alternative to commercial SPME fiber for the determination of accelerants in fire debris.

Keywords: Lab-made C₈-coated Fiber, HS-SPME, Arson, Accelerants, Fire Debris

1. Introduction

An important aspect of an investigation of a suspected arson case involves the chemical analysis of the debris remaining after the fire. Currently, accelerant extraction and analytical techniques have been refined to improve sample turnover and to reduce the number of inconclusive findings. For this purpose, headspace solid-phase microextraction (HS-SPME) has been introduced. The major advantage of this technique is that it uses no solvents and can be used for either direct sampling or sample clean-up. It is fairly economical and is a relatively simple and sensitive technique.

In SPME, the outer surface of a solid fused-silica fiber is coated with a selective stationary phase. Thermally stable polymeric materials that allow fast solute diffusion are commonly used as stationary phases. The extraction is based on the enrichment of components on a polymer or adsorbent coated fused silica fiber by exposing the fiber either directly to the sample or to its headspace. The qualities that enable an SPME adsorbent to be successfully used for accelerant extraction and analysis are its selectivity

towards accelerant components which separates and concentrates the accelerant from the headspace to yield a sample that is suitable for introduction to GC-FID [1].

Recently, many novel coatings have been developed using different techniques and technology for use in SPME. Compared with commercially available SPME adsorbents, the new materials exhibited longer lifetime (~150-200 times usage), higher thermal stability (<350 °C), higher extraction capability and a good solvent stability (organic and inorganic [2-4]). However, up to now, none of the novel fibers have been evaluated for the determination of accelerants in arson analysis. As a preliminary study, the HS-SPME procedure was optimized and evaluated using a lab-made SPME adsorbent comprising of sol-gel derived C₈-coating, with the aim of improving the quality of fire debris analysis.

2. Experimental

2.1. Chemicals and Materials

Individual standards of *n*-alkanes (C₈, C₁₀, C₁₂, C₁₄ and C₁₆) were purchased from Fluka Chemika. Samples of diesel and unleaded gasoline were purchased from a petrol station in Skudai, Johor while kerosene was obtained from a grocery shop at Taman Universiti, Skudai, Johor. Samples

¹Corresponding Author E-mail: umi@kimia.fs.utm.my,
Tel: +07-5534522, Fax: +07-5566162

of carpet were purchased from a carpet retail shop in Taman Ungku Tun Aminah, Skudai, Johor, Malaysia.

2.2. Apparatus

Two glass apparatus (400 cm³ and 125 cm³) for sample preparation step of HS-SPME was specially designed [5]. A Supelco SPME holder (Bellefonte, Pennsylvania, U.S.A.) and a used SPME fibre with a burnt off tip, coated with sol-gel derived C₈-coated fiber containing (1:1, C₈-TEOS:MTMOS) [6] were employed for the extraction of *n*-alkanes and accelerants. An SGT capillary column cutter with a rotating diamond blade from Shortix, Holland was used to cut the coated fibers for electron microscopy experiments.

2.3. Instrumentation

Gas chromatography analyses were conducted using a Hewlett-Packard 6890 GC (Wilmington, Delaware, U.S.A.). The HP 6890 gas chromatograph was equipped with FID and a HP ChemStation for data processing. An Ultra-1 capillary column (Agilent) of dimensions 25 m x 0.20 mm x 0.11 μm film thickness was used. Helium was used as the carrier gas at a flow rate of 1.2 mL/min. The injection port temperature was set at 250 °C and FID temperature at 310 °C. SPME injections were performed using a split mode injection (5:1).

Studies of fiber coating structure and thickness were made by means of Philips Scanning Electron Microscope model XL 30 SEM (Philips Electronic Instruments Company, Mahwah, New Jersey) equipped with a ThermoNoran energy dispersive X-ray detection system (EDX).

2.4. Procedure for Extractions using HS-SPME

In order to extract standard hydrocarbon compounds using the lab-made C₈-coated fiber, 30 μL from the prepared standard solution of *n*-alkanes (C₈, C₁₀, C₁₂, C₁₄, and C₁₆) was placed in the sample preparation apparatus which was immersed in a hot water bath and heated for 20 min at 100 °C. The C₈-coated fiber was exposed in the headspace for an appropriate time. Finally the exposed fiber was retracted and the fiber extracts were analyzed using GC-FID. The oven temperature was initially set at 40 °C, programmed at a rate of 10 °C/min until a final temperature of 270 °C.

2.5. Preparation of Fire Debris sample

A sample of carpet (20 cm x 13 cm) placed on a sheet of aluminium foil was ignited with a fire starter and left to burn until about one-third remained on the aluminium foil. Fire was extinguished by cutting off the oxygen supply. The partially burnt carpet was then exposed to the surrounding air for 30 minutes to let it cool down.

2.6. HS-SPME Optimization and Evaluation Procedures

2.6.1. Optimization of HS-SPME Operating Conditions

30 μL from the prepared standard solution of *n*-alkanes (C₈, C₁₀, C₁₂, C₁₄, and C₁₆) was placed in the sample preparation apparatus and followed by HS-SPME extraction. When the desired fiber exposure time was reached, the fiber was retracted and the analytes were thermally desorbed in the GC injection port. Several variables such as fiber exposure time, desorption time and extraction temperature were optimized by employing the procedures mentioned above.

The fiber exposure time was preliminary optimized by varying from 5-35 mins while keeping the extraction temperature (80 °C) and desorption time (80 sec) constant. The analyte desorption time at the GC injection port was then optimized from 10-90 sec by keeping the extraction temperature (80 °C) and extraction time (15 min) constant. Finally the extraction temperature was optimized by varying from 40-100 °C while keeping the fiber exposure time (15 min) and desorption time (80 sec) constant.

2.6.2. Determination of Limits of Detection (LOD)

The detection limits of fire accelerants were investigated using the lab-made C₈-coated fiber. Individual accelerants (petrol, kerosene and diesel) were spiked onto a blank burnt carpet sample in decreasing amount (initially 1 μL). Similar fire debris sample preparation as described in Section 2.5 was employed. The HS-SPME extractions were carried out with the optimized conditions of the C₈-coated fiber.

2.7. Scanning Electron Microscopy

Scanning Electron Microscopy (SEM) analysis was carried out in order to investigate the structure of the coated surface, estimate the thickness of the coating and determine the reproducibility of the coating thickness. The 1 cm segment of the coated fiber end was cut using a diamond cutter. The fiber was then placed on an aluminium stub with the help of a double edged carbon tape. The stubs were placed in a dessicator prior to SEM analysis. The SEM was carried out by the bombardment of electrons of 30 KeV on the target sample (C₈-coated fiber).

3. Results and Discussion

3.1. Optimization of HS-SPME Operating Conditions

3.1.1. Optimization of Extraction Time

The extraction equilibration time was initially investigated by exposing the fiber to the headspace of an apparatus containing target analytes for a variety of time, from 5 to 35 minutes, until the amounts extracted remains constant. The extraction time profiles were constructed by plotting the mass adsorbed, measured as chromatographic peak area, against extraction time as shown in Figure 1.

The extraction efficiency reaches a plateau after 25 minutes exposure of the fiber to the headspace. Therefore it can be suggested that after 25 minutes, an equilibration has been established. Within the duration of 25 minutes, the analytes would have migrated in and out of the coating until the equilibration was established between the phases.

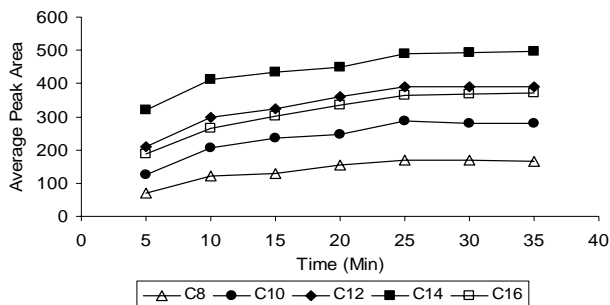


Figure 1. Extraction equilibration profiles of n-alkane hydrocarbon compounds (C₈, C₁₀, C₁₂, C₁₄ and C₁₆) at varying extraction time. Extraction conditions: extraction temperature of 80 °C and desorption time of 80 s.

According to a previous work done in this lab, Yong [5] suggested 35 minutes equilibration for components commonly found in accelerants using commercial PDMS fiber (cooling temperature ramp approach). A slightly shorter equilibration time (25 minutes) was obtained in this study by using the lab-made C₈-coated fiber as compared to those reported by Yong [5]. This might be due to the porous structure and thinner coating of the C₈-coated fiber. Lopes *et al.* [3] reported that coatings provided by sol-gel process are thinner and therefore saturation with the extracted analytes takes less time than with conventional fibers, that are coated with relatively thicker films of sorbents.

3.1.2. Optimization of Desorption Time

The time needed for complete desorption of analytes from the fiber was carefully determined to avoid sample carry-over problem. Nine desorption times ranging from 10 to 90 s in 10s increments were tested. The analytes were thermally desorbed at the GC injection port set at a temperature of 250 °C, at which the volatile hydrocarbon compounds can be desorbed immediately.

As can be seen from Figure 2 (relationship between response and desorption time), the chromatographic signal increased gradually with desorption time and stabilized at about 50 s, which indicated that a state of equilibrium had been reached. This result revealed that desorption process is fast and all the hydrocarbon compounds could be desorbed in 50 s. The optimized desorption time (50 s) can be considered as the minimum time for quantitative transfer of the analytes from the C₈-fiber to the GC column. However to ensure complete desorption, the desorption time was set at 1 minute.

The desorption equilibration time exhibited by the lab-made C₈-fiber is slightly shorter than those reported by Yong [5] using commercial PDMS fiber (cooling

temperature ramp approach). A short desorption time obtained in this study may be attributed to much larger surface area of C₈-coating which is favorable to rapid mass transfer from the porous and thinner coating. It has been reported that sol-gel coatings possess a porous structure that provides faster mass transfer during extraction as well as analyte desorption processes during sample introduction, which shortens the desorption time of analytes [7-9].

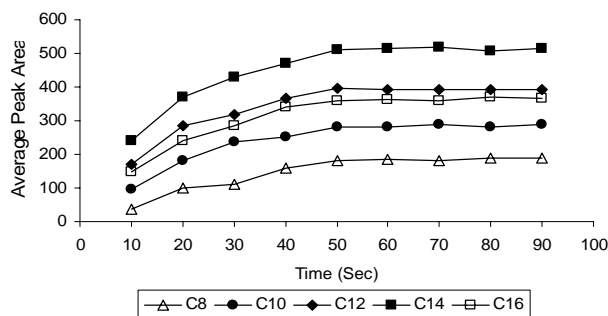


Figure 2. Desorption profiles of C₈-coated fiber for standard n-alkane hydrocarbon compounds (C₈, C₁₀, C₁₂, C₁₄ and C₁₆) by HS-SPME. Extraction conditions: extraction time, 15 min; extraction temperature, 80 °C.

3.1.3. Optimization of Extraction Temperature

The effect of extraction temperature on the amounts of analytes absorbed was studied by performing HS-SPME extractions of hydrocarbons at temperatures ranging from 40 to 100 °C. The average peak area versus extraction temperature profiles for HS-SPME extractions of standard hydrocarbons were determined as shown in Figure 3.

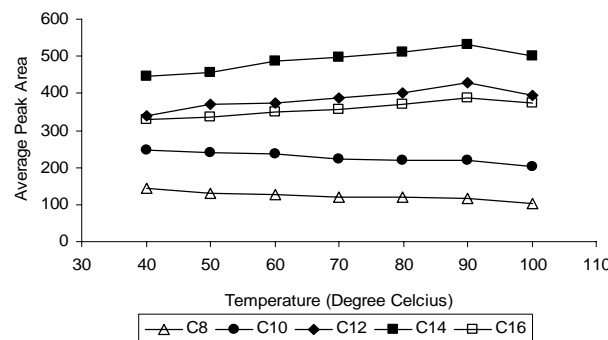


Figure 3. Influence of extraction temperature on the signal intensity of standard n-alkane hydrocarbons (C₈, C₁₀, C₁₂, C₁₄ and C₁₆). Extraction conditions: extraction time, 15 min; desorption time, 80 s.

It is apparent that peak areas increase as the extraction temperature is increased. The extraction efficiency of the fiber increased with temperature until 90 °C. Increasing the extraction temperature to 90 °C might have helped transfer of analytes to the headspace and enhanced diffusion of analytes towards the fiber. However, extraction yields tended to decrease slightly after 90 °C. This might be because the distribution constant decreased with increasing temperature. In order to control the stable

temperature efficiently, an extraction temperature of 90 °C was selected for analysis of volatile hydrocarbon compounds.

The extraction efficiency of sol-gel fiber is mainly controlled by two factors, distribution velocity and partition coefficient. The distribution velocity, which helps analytes to move near the solid-phase coating, increases with increased temperature. Extraction is an exothermic process and the partition coefficient, which determines the ratio of analytes extracted, is inversely related to temperature [7-9].

3.2. Method Validation

3.2.1. Accuracy of HS-SPME Method using C₈-coated Fiber

The accuracy of the developed method for three replicate analyses was determined by calculating the relative standard deviations (RSD) of the peak areas for all standard hydrocarbons (C₈, C₁₀, C₁₂, C₁₄ and C₁₆) in burnt carpet sample spiked at the same concentration level. In order to determine the reproducibility of the analytes peak areas, the three consecutive analyses were performed on the same day.

All the hydrocarbon compounds gave a low relative standard deviation (RSD) value ranging from 3.9 %-5.5 % which shows an acceptable reproducibility. These results indicated that quantitative analysis of hydrocarbons was possible using the external standard method and the applicability of lab-made C₈-coated fiber for routine analysis. The result obtained is in good agreement with those reported by Gbatu *et al.* [10]. It has been reported that the C₈-coated fiber exhibited good reproducibility for the analysis of organometals.

3.2.2. Detection Limits of Accelerants

Detection limits of accelerants (lowest detectable amount determined based on three times the baseline noise) were evaluated in order to access the performance of the proposed method. The lab-made C₈-coated fiber was employed for extractions of gasoline, kerosene and diesel in spiked burnt carpet sample under optimum HS-SPME conditions.

Referring to Table 1, for all the common accelerants, the LODs were between 0.7 to 1.0 µL. These values are indicative of good performance of the lab-made C₈-coated fiber. The detection limits of common accelerants obtained in this study using C₈-fiber is comparable to those reported by Yong [5] using commercial PDMS/DVB fiber.

Table 1. LODs for accelerants determined using C₈-coated fiber under optimum HS-SPME conditions

Common accelerant	Minimum detectable amount (µL)	
	This work	Previous work *
Gasoline	0.7	0.5
Kerosene	1.0	1.0
Diesel	1.0	1.0

* Yong [5]

3.3. Scanning Electron Microscopy (SEM) Analysis

3.3.1. Surface Characteristics of the Coating

The morphology of the lab-made fiber was investigated using the scanning electron microscope. As can be seen from Figure 4, it is obvious that the surface of the fiber coating was well-distributed. A homogeneous porous structure was suggested for sol-gel C₈-coating (Figure 5). Chong *et al.* [2] and Yu *et al.* [7] have reported that porous structure of the sol-gel coating should significantly increase the surface area on the fiber. Consequently, with such a porous coating structure, even an apparently thinner coating will be able to provide enhanced stationary-phase loadings and therefore, high fiber sample capacity. The porous structure of sol-gel coating helps faster mass transfer during extraction and therefore the equilibration time is shorter [8]. Figure 6 reveals that the fiber was not uniformly coated. This suggestion is agreeable to those reported by Gbatu *et al.* [10].

Examination of the scanning electron micrograph of the fiber prepared from 1:1 molar ratio of C₈-TEOS:MTMOS showed no apparent cracks (Figure 4). This is in good agreement with the previous work done by Gbatu. Works done by Gbatu *et al.* [10], showed that no apparent crack was visible for the fiber prepared from 1:1 molar ratio of C₈-TEOS:MTMOS but the cracks were visible in the fiber coated with a higher molar ratio (2:1). This could be due to the presence of the bulky C₈-TEOS group in a very large amount. Therefore all the analyses in this study were performed using fibers prepared from a 1:1 molar ratio.

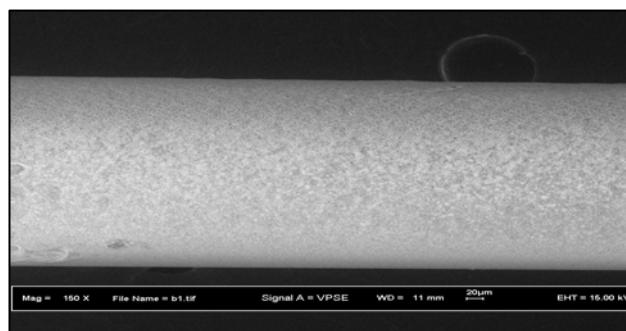


Figure 4. The surface view of C₈ fiber at 150 fold-magnification obtained by SEM.

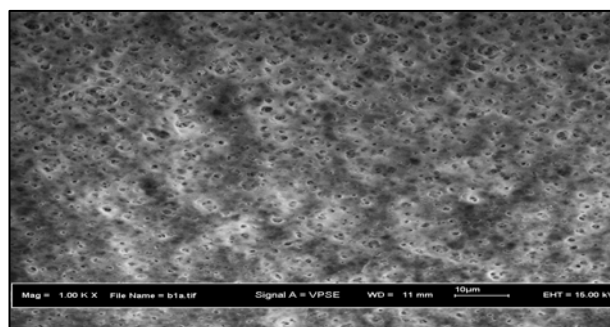


Figure 5. The surface view of C₈ fiber at 150 fold-magnification obtained by SEM.

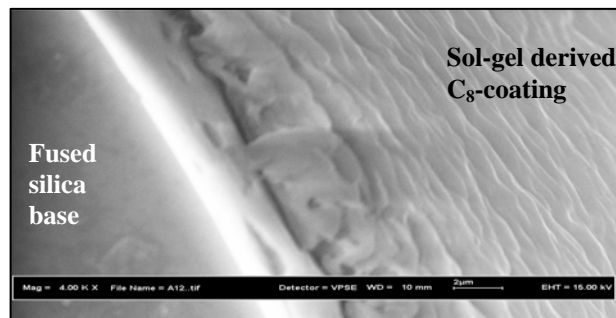


Figure 6. The cross-sectional view of C₈ fiber at 4000 fold-magnification obtained by SEM.

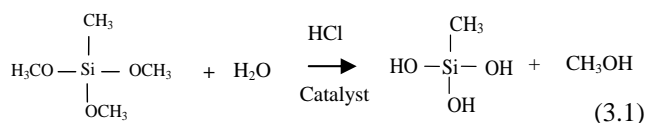
3.3.2 Estimation and Reproducibility of the Film Thickness

The film thickness of five different lab-made C₈-coated fibers were measured in order to estimate the thickness of C₈ layer and determine the reproducibility of the coating thickness. The thickness of the coating varied along the length of the fiber. It has been reported that the varied thickness obtained might be due to the rate at which the fused silica tip is removed from sol-gel solution after coating is not controlled and the portions of activated fused silica fibers that are exposed to the sol-gel solution for coating are not uniform [10]. Therefore the film thickness of each fiber was measured at two different points and the average coating thickness were calculated.

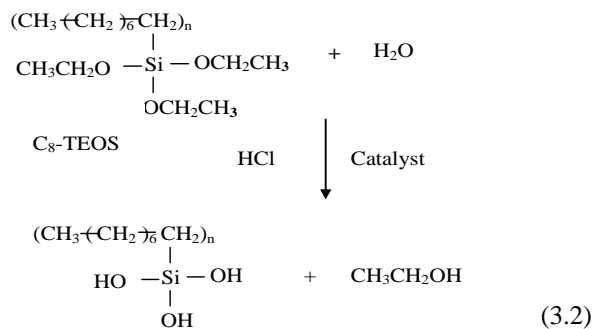
The average thickness of the C₈ coating was estimated as being ~ 3-4 μm, which is considerably thinner than most conventional SPME fibers (65 μm for PDMS/DVB). The reproducibility of the coating thickness obtained in this study is rather promising. Works done by Azenha *et al.* [11] revealed that the film thickness reproducibility was unsatisfactory for sol-gel PTMOS fiber.

3.4. Possible Underlying Mechanism of the Coating Process

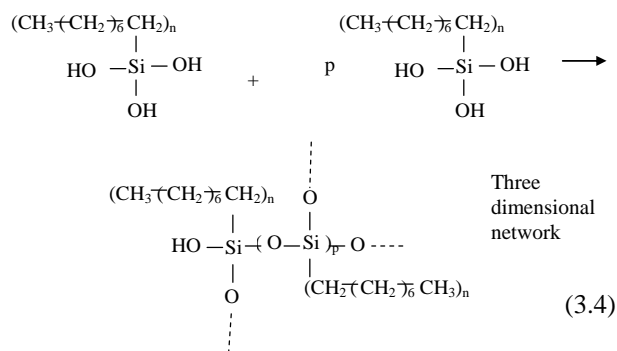
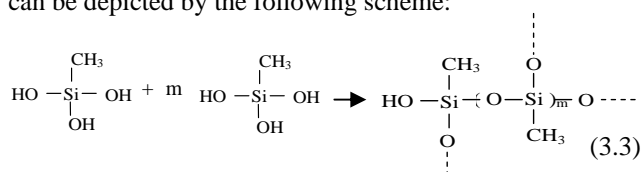
The in situ creation of the sol-gel C₈ coating involved two major sets of reactions [2-4, 7-9]: (1) hydrolysis of the precursor and (2) polycondensation of the hydrolyzed products. These reactions are catalyzed by acids or bases and lead to the formation of a polymeric network. In this study, MTMOS was used as a precursor and HCl as the catalyst. There are several steps during the sol-gel processing. The first step is the hydrolysis reaction of the precursor and co-precursor under the acid catalyst. The reaction can be presented by the following equation:



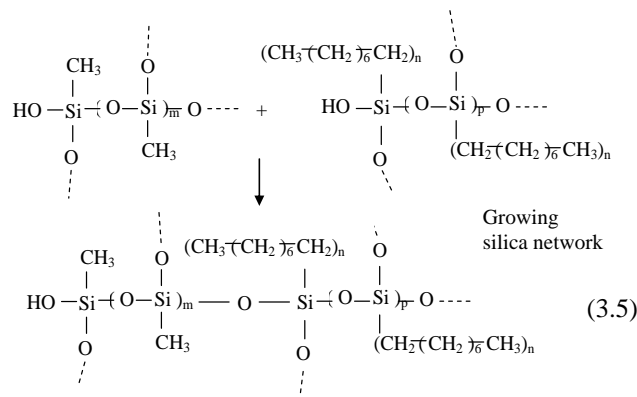
In the sol gel solution, C₈-TEOS is the coating stationary phase. However it also reacts as a co-precursor which contains hydrophobic alkyl chain (octyl group). The hydrolysis reaction of C₈-TEOS co-precursor can be presented by the following equation:



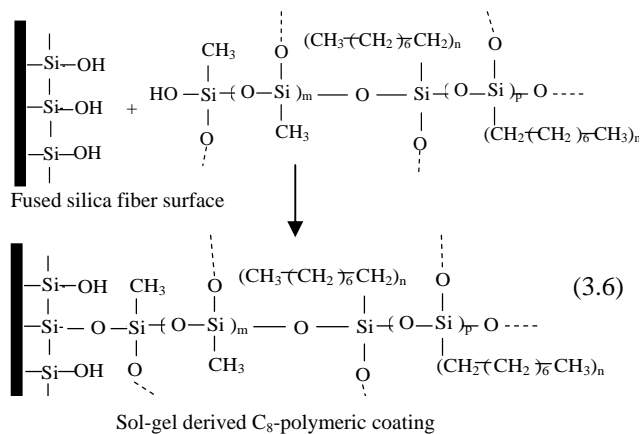
The second step is that the hydrolyzed products undergo polycondensation reactions to produce a three-dimensional polymer network. The octyl groups of the C₈-TEOS co-precursor do not participate in the polycondensation reactions, they are exposed and serve as the hydrophobic extracting phase. Polycondensation reactions can be depicted by the following scheme:



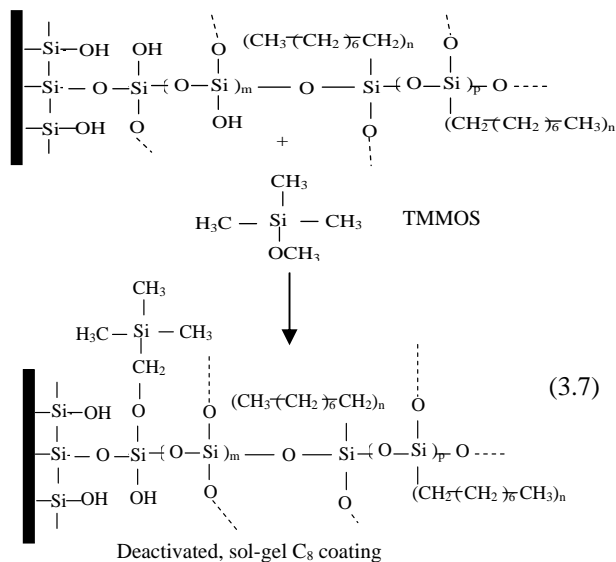
These reactions are followed by chemical binding of the coating stationary phase to the growing silica network. The chemical reaction involved can be schematically represented by the following equation.



The silanol groups on the fused-silica fiber surface can also join in the condensation reactions and provide chemical anchorage to the polymeric network in the immediate vicinity of the fiber surface (evolving sol-gel coatings). Schematically this reaction can be represented as follows:



The coated fibers were treated with a solution of TMMOS to reduce the silanol contents of the coating. Deactivation of surface-bonded sol-gel C₈-coating with TMMOS can be represented by the following scheme:



3. Conclusions

The headspace SPME method using lab-made C₈-coated fiber has been successfully optimized. A short extraction (25 minutes) and desorption (50 s) equilibration times arised from the porous structure of the sol-gel C₈-coated fiber and resulted in short analysis time. The optimum extraction temperature for standard hydrocarbons was found to be 90 °C. The developed HS-SPME method using C₈-coated fiber exhibited a high degree of inter-day accuracy under the optimum experimental conditions. The limits of detection for the common accelerants studied under the optimum HS-SPME conditions were in the range of 0.7 to 1.0 µL. Electron microscopy experiments revealed that the surface of the fiber coating was well-distributed but not uniformly coated and a porous structure was suggested for the sol-gel derived C₈ coating with an approximate thickness of (3-4) µm. The reproducibility of the film thickness was found to be rather promising. The resulting procedure was shown to be a good alternative methodology for arson analysis.

Acknowledgements

We would like to express our deepest gratitude to the Ministry of Science, Technology and Innovation, Malaysia (MOSTI) for the financial support under IRPA vote No. 74090.

References

- [1] Lord, H. and Pawliszyn, J. 2000. Evolution of Solid phase Microextraction Technology. *J. Chromatogr. A.* 885: 153-193.
- [2] Chong, S. L., Wang, D., Hayes, J. D., Wilhite, B. W. and Malik, A. 1997. Sol-gel Coating Technology for the Preparation of Solid-phase Microextraction Fibers of Enhanced Thermal Stability. *Anal Chem.* 69: 3889-3898.
- [3] Lopes, A. L. and Augusto, F. 2004. Preparation and Characterization of Polydimethylsiloxane / Poly(vinylalcohol) Coated SPME Fibers using Sol-Gel Technology. *J. Chromatogr. A.* 1056: 13-19.
- [4] Li, X., Zeng, Z. and Zhou, J. 2004. High Thermal-Stable Sol-gel Coated Calix[4]arene Fiber for SPME of Chlorophenols. *Anal Chim Acta.* 509: 27-37.
- [5] Yong, Y. T. 2004. Development of Headspace SPME-GC Technique for the Forensic Analysis of Ignitable Liquid Residues in Fire Debris. M. Sc. Thesis. Universiti Teknologi Malaysia.
- [6] U. K. Ahmad, A. R. Yacob and G. Selvaraju. 2005. A New SPME Adsorbent for the Forensic Analysis of Accelerant Residues. Ibnu Sina Institute, Universiti Teknologi Malaysia: Annual Fundamental Science Seminar Proceeding.
- [7] Yu, J., Wu, C. and Xing, J. 2004. Development of New SPME Fibers By Sol-gel Technology for The Determination of Organophosphorus Pesticide Multiresidues in Food. *J. Chromatogr. A.* 1036: 101-111.
- [8] Wang, Z., Xiao, C., Wu, C. and Han, H. 2000. High-performance Polyethylene Glycol Coated Solid-phase Microextraction Fibers Using Sol-gel Technology. *J. Chromatogr. A.* 893: 157-168.
- [9] Zeng, Z., Qiu, W., Wei, X., Yang, M., Huang, Z. and Li, F. (2001). Solid-phase Microextraction of Monocyclic Aromatic Amines using Novel Fibers Coated with Crown Ether. *J. Chromatogr. A.* 934: 51-57.
- [10] Gbatu, T. P., Sutton, K. L. and Caruso, J. A. 1999. Development of New SPME Fiber by Sol-gel Technology for SPME-HPLC Determination of Organometals. *Anal Chim Acta.* 402: 67-79.
- [11] Azenha, M., Malheiro, C. and Silva, A. F. 2005. Ultrathin Phenyl-functionalized Solid-phase Microextraction Fiber Coating Developed by Sol-gel Deposition. *J. Chromatogr. A.* 1069: 163-172.

Principles of novel continuous ion exchange plate column for Heavy metal removal

Idral¹ and Saidi, H.^{1*}

¹Membrane Research Unit, Department of Chemical Engineering,
Universiti Teknologi Malaysia, Jln. Semarak 54100, Kuala Lumpur, Malaysia

Tel : +60-3-26154823, Fax : +60-3-26914427

Abstract

In this paper, a novel ion exchange plate column has been introduced, combining cationic and anionic ion exchange in one column a co-current three phase. The study was based on the lab. scale. The different operating conditions and wastewater treated quality were investigated. The benefit of a concept in which metal ions removal are doing in the separate plates aimed to be reaching the completed reaction and also the resin consumption can be reduced which caused the operational cost of ion exchange be high at the metal ions removal technologies previously.

Assembling the continuous ion exchange plate column will be intended to be promises technology in the wastewater treatment, in which low cost for the investment and operational. The percentage removal of metal ions is according by the DOE requirements.

Keywords: Continous ion exchange system; Heavy metal ions removal; Plate column; Percentage removal

1. Introduction

Many efforts have been done to control and prevent water pollution in environment, like emphasizing the need for continuous refinement of existing treatment technologies, recognition and development of promising emergent technologies.

Several treatment processes, such as pH adjustment, chemical oxidation and reduction, electrochemical processes, ion exchange processes, membrane separation and biological treatment are available for reducing the heavy metals content in industrial wastewater. Ion exchange is the typical method that has been chosen for heavy metals removal from industrial wastewater treatment since this method is a well-established process and has been used in industrial operations for a long time.

The existing ion exchange system in industry is still disadvantageous, although it has reached high conversion. The disadvantages are of the expensiveness, batch system, resin consumption, and the inapplicability to irreversible reaction.

The novel design of ion exchange system has a potential advantage to be applied in industry, because the design is able to cover the disadvantages of the existing ion exchange system. In addition, the novel design is designed to be applicable for continuous systems. The capability to operate

in a continuous system will address to the decrease of both the investment and operational cost.

2. The novel continuous ion exchange

This paper presents principles of a novel process continuous ion exchange system can be used for removing metal ions from the industrial wastewater. Both cationic ion exchange and anionic ion exchange are removed as simultaneously. The concentration of metal ions in the treated wastewater was lower than 5 mg/l. The technique not only omitted degasifier but also reduced the cost in construction and runtime of treated wastewater to discharge the another way.

3. Experimental set-ups

3.1. Continuous ion exchange and supporting equipment:

The experimental setup used in this study consisted of the main ion exchange column, regenerating column, tanks, pulsation system, pump and flow measuring devices. Figure below shows a schematic diagram of the setup. Material will be chosen is must be compatible with the solution used to avoid corrosion and solution contamination by corrosion products. In the following sections, each unit in the

experimental setup is described. Detailed drawings and dimensions will be given when applicable.

3.1.1. The column and resin

The column that will be used is made of glass with a wall thickness of 0.01 m. The inside diameter of the column is 0.10 m and its total height is 1.50m. Figure 1. shows a detailed drawing of the column with its different sections. The column is divided into 13 compartment by perforated plates (10 perforation of $\varnothing = 6$ mm). The liquid feed is from below, the resin feed at the top. Resin and liquid flows and the total resin hold up in the column are carefully measured. The column is the similar processing one used for batch experiments.

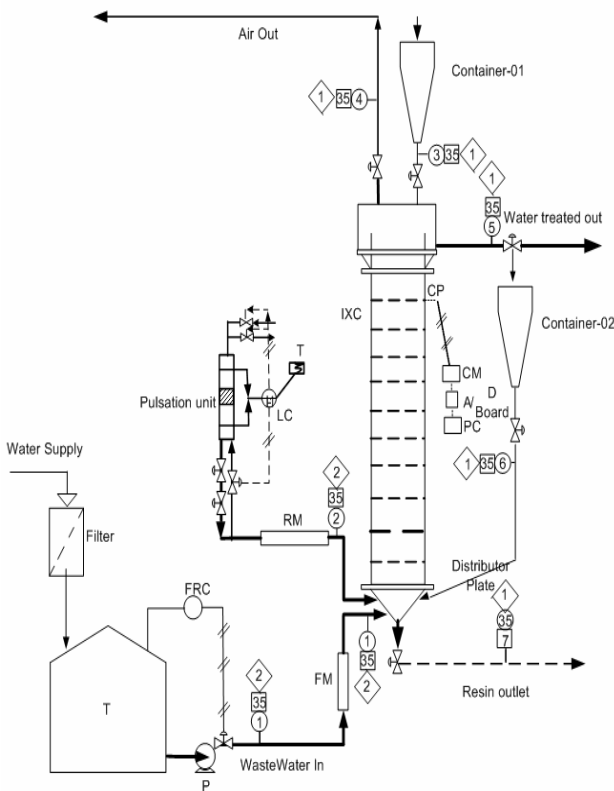


Fig.1. Schematic representation of the novel ions exchange plate column for metal removal from the industrial wastewater

3.1.2. Air flow course:

The air that will be used in the system is compressed air, which can be obtained from the compressor station. The pressure of the air is controlled using valve before rotary meter. This operating pressure is high enough to maintain the system can perform according by the design. Details of the system are given in the picture below.

3.1.3. Liquid flow course:

The liquid used in this study is heavy metal ions that representing of wastewater in the industry. The liquid is circulated continuously through ions exchange column.

The liquid is pumped from the wastewater tank using pump to the liquid distributor at the bottom of column. Its flow rate is measured using flow meter (FM).

The liquid leaving the column from the overflowing liquid from the top is measured with another flow meter (FM).

The pressure drop between the two ports of a plate is measured using an inclined liquid-filled manometer. A calibration must be performed first to obtain a relation between the manometer reading and the liquid flow rate under operating conditions.

3.2. Temperature variables

The variable of temperature is performed to find the effect of liquid temperature coming to column to the percentage metal ions removal. The measuring of temperature is only to do for the liquid temperature coming and leaving of the column.

3.3. Resin:

The resin used in this study is spherical resin with diameter 0.05 mm to 0.53 mm. The density of resins that will be used from the 1.04 g/ml to 1.27 g/ml, and also having exchange capacity around of the 0.8 moleq/ml to 3.5 moleq/ml.

3.4. Experimental measurements:

Several parameters will be measured in this study, namely the static pressure, the liquid solution conductivity, the heat liquid entering leaving of column, flowrate of liquid and air. These parameters is used in estimating many important operating characteristics for hydrodynamics, solid-liquid heat transfer, and solid liquid mass transfer.

The appropriate measuring technique is selected based on the following criteria: high accuracy, safety, easiness of use, and the ability to have the measurements done using a computerized data acquisition system. In the following sections, review of these varies techniques used are described.

3.4.1 Static pressure measurement

Static pressure is necessary to get several important characteristics such as be height and the phase hold-ups. A profile of the change in the static pressure is obtained by having several measuring points along of the column. These measuring points are connected to a bank of pressurized liquid-filled manometers.

Several experiments will be done in which the static bed heights is measured independently at mid bed condition. A specially designed inverse manometer is employed in this case.

3.4.1.1 Pressurized liquid-filled manometer

Pressurized liquid-filled manometers enabled the measurement of the static pressure accurately without the need to climb up ladder. The manometer, which will be made of glass tubing with I.D.=6mm, is connected to the pressure taps along the column wall using flexible polyethylene tubing. The liquid flowed from the column into the manometers. The other end of all manometers is connected to a header made of CPVC tube, I.D.=0.0127m (0.5"). The pressure in the header is increased using compressed air. The header pressure is recorded using a liquid-filled magnehelic pressure gage. The height of the liquid in the manometer tube is adjusted so that a clear and easy to read liquid column could be obtained.

The column have thirteen pressure measuring stations (taps). These taps are made so that the are flushed with the inside wall of the column. The distance (z) of these taps relative to the grid level is as follows, Z(m) =0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3

At any height, two taps are made in the column at right angles as shown in Figure 2. On side of these taps is used for pressure measure measurements, the other side is used for inserting the mass transfer probes in the figure.

Fluctuations in the pressure in the column could affect the manometers' reading. Therefore, a capillary will be inserted in the flexible polyethylene line to dampen the pressure fluctuations. The capillaries are made of Teflon with an inside diameter of 0.5mm and its length is 4-6 cm, Figure 2 shows the location of the capillary within the manometer pressure line. The capillary is installed at low point in the line to avoid trapped air bubbles from reaching the manometer.

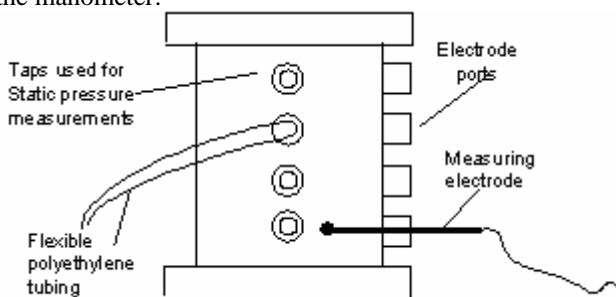


Figure 2. Schematic Diagram for a Section of the Column Showing Pressure Taps and Electrode Ports.

3.4.1.2 Inverse liquid-filled manometer

This manometer is used in several experiments to get a highly accurate measurement of the pressure drop in the ion exchange. Two pressure taps at 0.05m and 0.13m below the grid level are connected to the inverse liquid filled manometer using flexible tubing.

A capillary system similar to the one used with the regular pressurized liquid-filled manometers is installed on both lines. However, it is decided that it is not needed since the connecting lines and the inherent design characteristics of the inverse manometer eliminated any pressure fluctuations. Figure 3 shows a schematic of the inverse manometer.

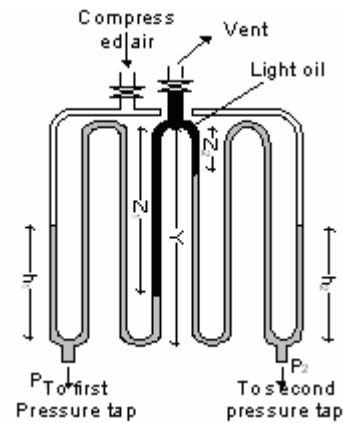


Figure 3. Schematic of the Inverse Liquid-Filled Manometer

The pressure exerted from the two pressure taps at the column is transferred to the bottom of the manometer so that a height difference in the liquid level is established. On the two outmost sides is an ordinary pressurized liquid filled manometer which would give a small difference Δh . The pressure difference is also exerted on the inverse oil-filled part of the manometer.

3.4.2 Electrical Conductivity Measurements:

Electrical conductivity measurements have been successfully used for phase hold up measurements in gas-liquid packed beds, (Achwal and stepanek, 1975, 1976), two-phase fluidized beds, (Turner, 1976) and three-phase fluidized bed, (Begovich and Watson, 1978; Ibrahim et. Al., 1996)

Stainless steel (SS304) tubes (I.D.=3mm) are used as the conductive electrodes as shown in Figure 4. The tubes are chosen over solid rods because simultaneous measurements of static pressure and conductivity are desired. The configuration shown in figure 4 shows how this objective are fulfilled.

The figure 4 shows a section in the column where the two stainless steel tubes, forming the conductivity electrodes, are inserted so that they are flushed with the column inside wall. One tube is connected to the manometer for static pressure measurements. The other is sealed at the end with a Silicon sealant. The two tubes are connected to a conductivity circuit which is designed so that the conductance of the solution in the column between the two electrodes could be sampled using a data acquisition system as shown in the figure 4.

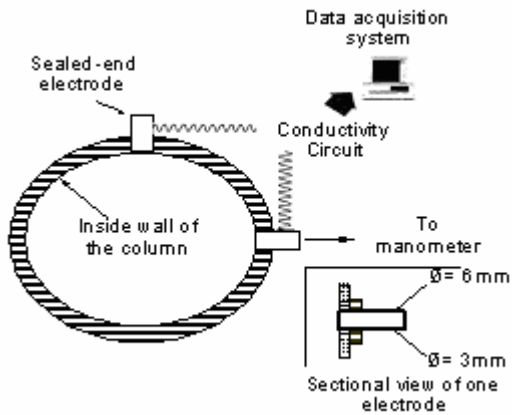


Figure 4. Electrical Conductivity Probes configuration and Connection

We will be employed six sets of conductivity electrodes, each set consisted of two stainless steel tubes and its own electric circuit. The location (distance below the grid) of these sets are as follows, $z(m) = 0.20, 0.40, 0.60, 0.80, 1.00, 1.20$. The objective of having these many sets is to be able to get the axial variation of the electrical conductance by sampling the six sets at the same time and within the same experiment and to do this at the shortest time possible.

The data acquisition will be performed using an Analog-to-Digital data acquisition card. A Quick Basic Program will be generated to perform the data acquisition, which used the PC-Lab.

3.4.3 Mass Transfer measurements

Solid-liquid mass transfer rates will be measured by the limiting current technique. This technique depends on the evolution characteristics of a diffusion controlled electrochemical reaction. Such technique has been used extensively for mass transfer measurement studies in different kinds of reactors, (Selman and Tobias, 1978). Several researchers have used this technique for mass transfer study in fluidized beds, (Del Pozo, 1994; Noire et. Al., 1992; Burru and Briens, 1991; Nikov and Delmas, 1987, 1992; Fukuma et. Al., 1988). This technique can be used to get the radial or axial distribution of the mass transfer coefficient in the column.

In the following sections, the principle of the electrochemical technique for measuring mass transfer will be discussed. The discussion will also cover the electrochemical system used, the mass transfer circuit, design of probes.

3.4.3.1 Principle of the method:

The electrochemical (limiting current) technique is based on the use of the Plateaux region of the curve in which the current passing through the cell does not depend on the applied potential. Instead, the current will depend on the mechanics of the system and how the ions are diffusing to the cathode. The rate with which the ions are arriving at the cathode is given by Faraday’s law as follows,

$$S_i I = - n F N_i \dots\dots\dots(1)$$

Where:

- F : Faraday’s constant, 96500 C/mol.
- I : current intensity, A/m²
- n : number of electrons
- N_i : molar flux of component I, mol/m².s
- S_i : number of molecules or ions of species i participating in the transfer n electrons

The molar flux can be found from the general mass transfer equation,

$$N_i = k_{ls} (C_i - C_b) \dots\dots\dots(2)$$

Where:

- k_{ls} = the mass transfer coefficient m/s
- C_i = concentration of the ions at the interface, mol/m³
- C_b = concentration of the ions in the bulk, mol/m³

Under limiting current conditions, the concentration at the interface (C_i) is virtually zero. Therefore, the mass transfer coefficient can be found by measuring the current and the concentration in the bulk of the solution.

Accordingly, the total ions reaching the cathode will be higher than the ions reaching by diffusion alone. Thus, the mass flux will be given by the following equation (Lin et. Al., 1951).

$$N_i = -D_{eff} \frac{\partial C}{\partial x} - C_u \frac{\partial \Phi}{\partial x} \dots\dots\dots(3)$$

where:

- D_{eff} : effective diffusivity, m²/s
- X : distance away from the electrode, m
- U : ion mobility, (m/s)/(V/m)
- Φ : Potential, V

Based on the above discussion, Equation (1) through (3) reduce to the following equation under limiting current condition,

$$k_{ls} = \frac{I_{lim}}{nF \cdot C_b \cdot A_e} \dots\dots\dots(4)$$

- where: I_{lim} : limiting current,
- A and A_e : surface area of the electrode, m²

3.4.3.2 Electrode design:

Several cathodes is inserted into the ion exchange column at different heights to get the axial profile of the mass transfer coefficient. They are also moved horizontally to obtain local values for the mass transfer coefficient at nine different radial locations similar to the solid-liquid heat transfer experiments.

The anode used is a 10 cm by 10 cm square plate made of silver, 1 mm thick. The plate is fixed firmly to the inside of the column and is connected to the electric circuit. The location of the plate is found to have negligible effect on the results, (Del Pozo, 1992), and is, therefore, fixed at 1.10 m

below the grid. The calomel electrode is connected to the ion exchange column using a flexible tubing at a location 0.25m below the grid.

An electric circuit is used to apply the appropriate potential to the cathodes and get the cell current for each electrode.

3.4.3.3 Electrode activation:

The polarographic wave or plateau shown in Figure 5 can only be obtained if the mass transfer electrodes are at their peak performance. Berger and ziai indicated that generally speaking activation can be neglected at low Reynolds number < 10⁴, whereas at higher Reynolds numbers, cathodic activation is a must.

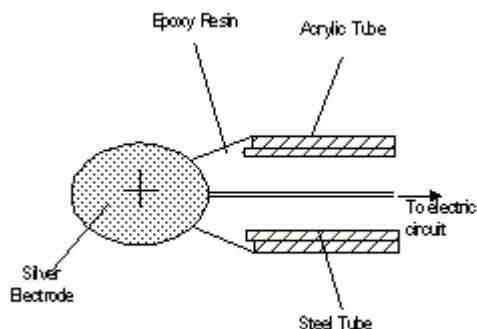


Figure 5 Design of the silver chatode

In this study, the following precautions are taken,

- The column is not emptied from the solution during the whole course of experimentation. In case the column had to be drained, it is washed thoroughly with de-ionized water
- The stability of the electrode signal is checked every day before any experimentation. Any drop in the cell current would be indicated electrode deactivation.
- An activation procedure including surface cleaning and cathodic activation is employed whenever needed to activate the electrodes.

3.5. Terminal Velocity Measurements:

The terminal velocity is an important parameter that is often used in the modeling of the hydrodynamics of ion exchange column, especially the bed expansion. Terminal rise velocity of light particles in liquids may differ from the terminal settling velocity of heavier particles in liquids, (Karamanev and Nikolov, 1992a). This difference has been attributed to the difference in the drag coefficient for light particles from the drag coefficient on heavier particles. Standard drag curves are then unsuitable to be used for light particles.

The measurements will be performed in the phosfat column, as shown in Figure 6. A 0.5 inch phosfat pipe is fitted to the base of the column at the centre. The particle motion is filmed with a video camera connected to a time coding device. The time coding device allowed measuring

the time in 1/30 of a second with a possibility of identifying 60 frames in one second. Using the video film, we are able to find the time each particle took to travel a certain axial distance marked on the side of the column.

The particles are filmed at three different sections along the column, at 10cm, 30cm, and 130cm above the entrance.

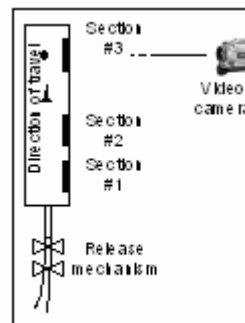


Figure 6. Terminal Velocity Measurements Assembly.

4. Experimental results

Bases on the principles of the continuous ion exchange plate column design and results simulation, we found the kls is as function of metal ions concentration in resin is as (kls)=1.864 / (16.836 + QA(i)) and effective diffusivities (De) = 1.116E-06 cm²/second

5. Conclusion and discussion

With understanding the principles of the design ion exchanges system for metal ions removal, we can determine significant variables that needed is as consideration before decided the variable design of ion exchanges. So that time and cost of process development can be significantly reduced through principles of the use of state-of-art process modeling together with lab experiments and miniplant validation. Using this combination it is often possible to significantly reduce the need of the traditional pilot-plant stage in the development of a new process.

References

- [1] Achwal, S.K. Stepanek, J.B. 1975."An alternative Method of Determining Hold-up in Gas Liquid Systems, Chemical.Eng. Sci., 30(11), 1443-1444
- [2] Achwal, S.K..Stepanek, J.B. 1976 "Hold up profile in Pacjed Beds," Chemical.Eng. Journal, 12(1),69-75
- [3] Turner, J.C.R.1976."Two-Phase Conductivity: The electrical Conductance of liquid-Fluidized beds of Spheres," Chem. Eng. Sci., 31(6), 487-492
- [4] Begovic,J.M.Watson,J.S.1978a."AnElectroconductivity Technique for the Measurement of Axial Variation of Holdups in Three-Phase Fluidized Beds,"AIChE Journal, 24(2), 351-354

- [5] Begovich, J.M., 1978b. "Hydrodynamic Characteristics of Three-Phase Fluidized Beds" in "Fluidization, proceedings of the 2nd. Engineering Foundation Conference, editors: Davidson, J.F., Keairns, D.L., 190-195, Cambridge, Cambridge University Press.
- [6] Ibrahim, Y.A.A., Briens, C.L., Margaritis, A., Bergougnou, M.A. 1996. "Hydrodynamic Characteristics of a Three-Phase Inverse Fluidized-Bed Column, AICHE Journal, 42(7), 1889-1900
- [7] Seleman, J.R., Tobias, C.W. 1978. "Mass Transfer Measurements by the Limiting-Current Technique," in "Advances in Chemical Engineering," edited by: Drew, T.B.; Gokalet, G.R., Hoopes, J.W.Jr., Vermeulen, T. Volume 10, 211-319, Academic Press
- [8] Del Pozo, M., Briens, C.L., Wild, G. 1994. "Effect of liquid coalescing properties on mass transfer, heat transfer and hydrodynamics in a three-phase fluidized bed," Chem. Eng. Journal, 55(1), 1-14
- [9] Noire, I., Briens, C., Margaritis, A., Wilg, G., 1992, « Hydrodynamics, Gas-Liquid Mass Transfer and Particle-Liquid Heat and Mass Transfer in Three-Phase Fluidized Bed for Biochemical Process Applications, » Chem. Eng. Sci., 47(13/14), 3573-3580
- [10] Burru, I.G., Briens, C.L., 1991a, "Particle – liquid Mass Transfer in Three-Phase Fluidized Beds. Part I. Newtonian Liquid," Powder Technology, 68(3), 243-253
- [11] Burru, I.G., Briens, C.L., 1991b, "Particle – liquid Mass Transfer in Three-Phase Fluidized Beds. Part II. Non-Newtonian Liquid," Powder Technology, 68(3), 255-261
- [12] Nikov, I., Delmas, H., 1992. "Mechanism of Liquid-Solid Mass Transfer and Shear Stress in Three-Phase Fluidized Beds," Chem. Eng. Sci., 47(3), 673-681
- [13] Nikov, I., Delmas, H., 1987, "Solid-Liquid Mass Transfer In Three-Phase Fixed And Fluidized Beds," Chem. Eng. Sci., 42(5), 1089-1093
- [14] Fukuma, M., Sato, M., Muroyama, K., Yasunishi, A., 1988, "Particle to Liquid Mass Transfer in Gas-Liquid-Solid Fluidization," Journal of Chemical Engineering of Japan, 21(3), 231-237
- [15] Lin, C.S., Denton, E.B., Gaskill, H.S., Putnam, G.L. 1951, "Diffusion-Controlled Electrode Reactions, Ind. Eng. Chem. 43(9), 2136-2143
- [16] Delpozo, M., 1992, "Transfer de Matiere et de Chaleur Particle-Liquide sn lit Fluides Gas-Liquid-Solide," Ph.D. Thesis, INPL, Nancy, France.
- [17] Karamanev, D.G., Nikolov, N.L. 1992a, "Free Rising Spheres Do Not Obey Newton's Law for Free Settling," AICHE Journal, 38(11), 1843-1846.
- [18] Berger, F.P., Ziai, A., 1983, » Optimization of Experimental Conditions for Electrochemical Mass Transfer Measurements, » Chem. Eng. Res. Des., 61(6), 377-382

Preparation and Characterization of Delaminated Zeolite, ITQ-6

Noor Aishikin M. Y. and Zainab R.

Department of Chemistry, Faculty of Science,
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia.
Tel: +60-75534491, Fax: +60-7-5566162, E-mail: zainab@kimia.fs.utm.my

Abstract

Recent discovery of the delaminated zeolite in the zeolite family has opened up new possibilities for preparing catalysts with larger access of large molecules to the active acid sites. Despite extensive research, few mesoporous zeolitic materials with strong acid sites have been successfully obtained. In this study, modification of PREFER, a kind of microporous zeolite having ferrierite structure to mesoporous material, ITQ-6 was carried out. The modification was done by delamination technique. The starting material, PREFER (precursor of ferrierite) was first synthesized in which after calcination resulted in the formation of microporous ferrierite (FER) type material. Modification to mesopore material by delamination of PREFER was performed using cetyltrimethylammonium bromide, CTABr as swelling agent, followed by sonication and calcination to form ITQ-6. The XRD results show that the crystalline phase of PREFER diminished for sample after delamination. Surface porosity study of the ITQ-6 sample has shown successful formation of homogeneous mesopores in size between 3-4 nm with higher external surface area. The acidity study using pyridine as a probe molecule and monitored by IR, indicates that the ITQ-6 is still having the superacidity property of zeolite. The FESEM micrograph shows that the plate like crystal of PREFER was destroyed upon delamination to give amorphous form of ITQ-6. The resulting ITQ-6 material with both acidity and mesoporosity can become a potential catalyst for bulky molecules.

Keywords:

Delaminated zeolite and ITQ-6.

1. Introduction

Throughout the 1990s, environmentalism remains a foremost concern, with zeolite in the forefront as solutions to new challenges in the generation of "Green Chemistry". The versatility of zeolite has allowed its application particularly as heterogeneous catalyst in acid [1], base [2] and redox [3] reactions. Over a wide range of solid acids, zeolites hold high acidities on their surface and have received much attention in the petroleum industry [4] and in organic synthesis [5].

Zeolites are significantly more active than the layered structures (clays) and mesoporous structures (whose walls are amorphous), and they impart shape selectivity on the reaction products. However zeolite presents some limitations when large reactant molecules are involved, especially in liquid phase systems, as is frequently the case in the synthesis of fine chemicals [6]. Providing access for larger molecules to the catalytic sites would expand the range of reactions that can be catalysed by zeolite. Attempts to improve the catalytic sites have so far focused on increasing the zeolite pore sizes [7] as well as decreasing the zeolite crystal sizes [8]. But attempts to increase the pore size of zeolites have met with only limited success. Pioneer works by Corma *et al.* [9] have show that delamination of layered zeolite precursor would gives a new zeolitic material, a kind of aluminosilicate whose

zeolite-type catalytic sites are contained within thin, readily accessible sheets. The synthesis of a partially crystalline bimodal pore system with combined micro and meso pores, designated as ITQ, has been achieved by delaminating the layered zeolite MCM-22. The delaminated zeolite, ITQ-2 first introduced by Corma *et al.* [9] presents a very high external surface area ($>600 \text{ m}^2 \text{ g}^{-1}$) with better accessibility to active sites for bulky molecules of interest in oil refining, chemicals, and fine chemicals compared to microporous zeolite. Delaminated zeolite is also stable towards high-temperature calcination.

The delaminated zeolite was proved to be active in the acetalization of phenylacetaldehyde with glycerol [10], epoxidation of 1-hexene [11] and hydroxyalkylation of 2-methoxynaphthalene with paraformaldehyde [12].

Besides ITQ-2; ITQ-6 [13], ITQ-18 [14] and LRS-1 [15] had been successfully been prepared based on the same delamination concept. Therefore this paper reports the preparation and some physicochemical characterization of the delaminated material known, as ITQ-6.

2. Methods

2.1 Materials

The PREFER material was synthesized with the ratio of $10\text{SiO}_2:1\text{Al}_2\text{O}_3:15\text{NH}_4\text{F}:5\text{HF}:10\text{R}:100\text{H}_2\text{O}$ where R is the template (R=4-amino-2,2,6,6-tetramethylpiperidine) used following the method described by Schreyeck *et al.*

[16]. The as-synthesised material, denoted as PREFER (the precursor of ferrierite) were kept uncalcined in order to preserve the template between the layers for further modifications.

ITQ-6 sample was prepared by swelling the laminar PREFER according to the procedure described by Concepcion *et al.* [17]. The PREFER layers was swollen by intercalation of PREFER with the solution of CTABr (25wt%), TPAOH (40wt%) and distilled water. The suspension obtained was refluxed and stirred vigorously for 16 hours at 80 °C. Finally the suspension was filtered, washed and dried. This bulked laminar material (swollen of PREFER layers) was identified as preITQ-6. The completion of the swollen material was monitored using X-ray diffraction.

Deionized water was added to the bulked laminar materials, preITQ-6 obtained before and this suspension obtained after vigorous agitation was subjected to ultrasonic treatment to force the layers apart. Thereafter the solid sample was recovered by filtration followed by drying at 100 °C overnight and this material was assigned as ITQ-6-bc. The removal of the organic templates was done by calcinations of ITQ6-bc at 550 °C yielding ITQ-6.

A portion of PREFER material was calcined at 550 °C without previous treatment to yield the formation of microporous FER type zeolite.

2.2 Apparatus and procedure

The extent of crystallization and phase purity was evaluated for all samples by recording X-ray diffractograms using Seimens 500 Kristalloflex with Cu K α radiation, ($\lambda = 1.54056 \text{ \AA}$, 40 kV, 40 mA) in the 2θ range of 2-40° at ambient temperature. Scanning was carried out in the step interval 0.05° with counting time of 1 second per step.

Framework vibration, surface hydroxyl groups and acid sites were determined using Perkin Elmer Spectrum One FTIR spectrometer. ITQ-6 framework was characterized using KBr technique. On the other hand for surface hydroxyl group and acid sites study, self supported disc was prepared and placed in a glass cell equipped with CaF₂ windows. ITQ-6 sample was heated at 400 °C under vacuum condition ($P \sim 1 \times 10^{-7}$ mbar) overnight. Surface hydroxyl groups were recorded at room temperature. For acidity study, the sample was dehydrated at 400 °C under vacuum ($P \sim 1 \times 10^{-7}$ mbar). Adsorption of pyridine on dehydrated sample was carried out at room temperature followed by desorption under vacuum at 150 °C and 400 °C for 1 hour respectively in order to determine the strength of the acid sites at ambient temperature.

The specific surface area and the pore size distributions of ITQ-6 were determined by BET technique using ASAP 2010 instrument.

Field Emission Scanning Electron Microscope (FESEM), Carl Zeiss Supra Series model was used to determine the morphology of sample and crystal size of ITQ-6.

3. Results and discussion

3.1 X-ray Diffraction (XRD)

ITQ-6 can be obtained through delamination process of layered zeolite ferrierite precursor. The PREFER sample was first swollen to form bulk material by intercalation of big molecule, cetyltrimethylammonium bromide (CTABr) between the PREFER layers. This bulk material is assigned as preITQ-6. Then the preITQ-6 was delaminated through sonication to produce ITQ-6 followed by calcinations in order to remove the templates. This will lead to the formation of mesoporous ITQ-6 material. All the treatments to the PREFER sample was monitored using XRD technique.

The schematic process of preparing ITQ-6 is illustrated in Figure 1. Synthesis of delaminated ITQ-6 [14] material may be made via preparation of gel, its hydrothermal treatment, and the resulting material (PREFER) is treated with a swelling solution, an organic compound through ion exchange. This is followed by partial delamination, using by mechanical agitation or ultrasonic. The final oxide material obtained is calcined to remove the templates to give ITQ-6.

Delamination can be done in a similar way, which can be done with clays. This allows access of large molecules to active sites, which would normally have been precluded by the size constraints, imposed by the pore structure. This procedure makes the preparation of a new material possible and the result is a material called ITQ-6. It is formed by single layers organized in a "house of cards"-type structure, which is thermally stable and presents a well-defined and homogeneous external surface area

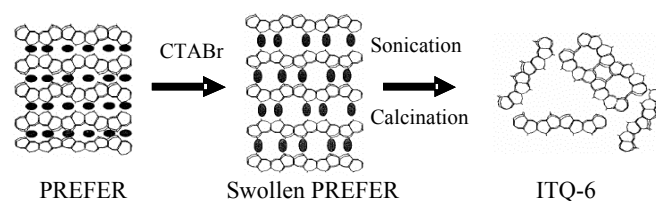


Figure 1: Scheme for preparing ITQ-6 material.

Figure 2 shows the X-ray diffractogram of PREFER samples after each treatment. As can be seen in Figure 2, the swollen material, preITQ-6 shows an increased in the d-basing of d_{200} plane in the range 1.31-3.00 nm compared to 1.30 nm for PREFER while the peaks corresponding to the individual layers decreased strongly. This indicates that the PREFER layers containing 4-amino-2,2,6,6-tetramethylpiperidine templates had been expanded in the present of CTABr as swelling agent. When the delamination was completed as in sample ITQ-6-bc, the XRD peaks of the sample are much broader and less intense than that of PREFER and preITQ-6. A sharp peak was observed at 2θ 6.8° indicating the presence of a small amount of PREFER that had not been swollen and delaminated.

Upon calcinations of the delaminated ITQ-6-bc sample, much broader and less intense peaks were observed in the X-ray diffractogram of ITQ-6 sample compared to the ITQ-6-bc sample. The X-ray diffractogram of ITQ-6 sample also shows a sharp peak at 2θ 9.4° which is due to the existence of FER type zeolite due to the removal of templates in the delaminated ITQ-6-bc sample that contain a small amount of PREFER.

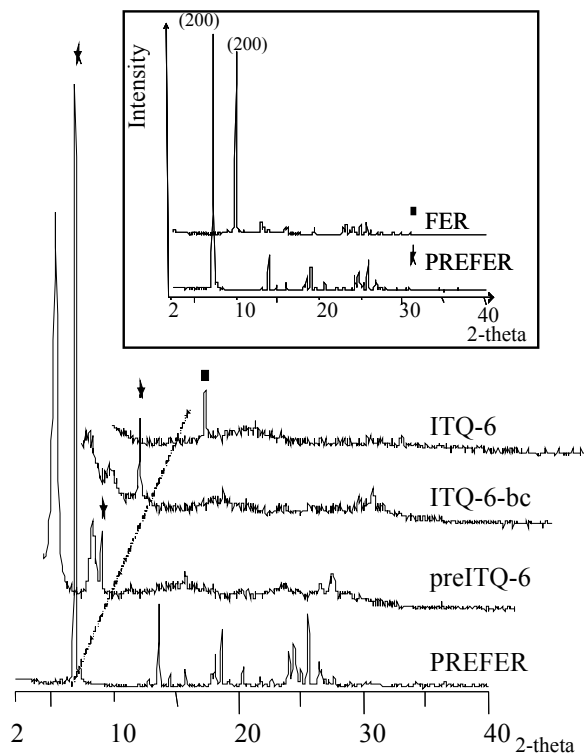


Figure 2 : X-ray diffractograms of PREFER samples after every treatments (insert shows XRD pattern of FER and PREFER)

Microporous FER type zeolite can be obtained through complete removal of template by calcination of PREFER material at 550°C . During calcinations of two-dimensional sheets of PREFER, the condensation of the silanol groups and elimination of the organic template occur; this accompanied by a decrease of the interlayer spacing to form three-dimensional FER type material.

X-ray diffractogram of PREFER material and FER type zeolite (calcined PREFER) is presented in Figure 2 (small box). The XRD patterns of both samples are practically identical and characteristic of highly crystalline products except for the peak at d_{200} .

Table 1 : Values of unit cell parameters for PREFER and FER type zeolite.

	a (Å)	b (Å)	c (Å)
PREFER	26.09	13.94	7.42
FER	18.68	13.99	7.49

As shown in Table 1 the values of the parameter *b* and *c* are unchanged, whereas the *a* parameter decreased

drastically upon removal of templates by calcinations. Finally all these observation suggest that the structural changes leading from the PREFER structure to the FER type structure involved essentially the *a* axis.

3.2 Fourier Transform Infrared (FTIR)

3.2.1 Framework

The IR spectra of the ITQ-6 framework and hydroxyl group are presented in Figure 3. The framework of ITQ-6 was found to be similar to the amorphous silica and in agreement with X-ray diffractogram findings. The band at 430 cm^{-1} assigned for pore structure of ferrierite framework diminished and the intensity of the characteristic peaks of ferrierite framework at 721 cm^{-1} and 580 cm^{-1} decreased upon delamination process. It can be observed from the IR spectra that the strongest band corresponding to asymmetric stretching around 1080 cm^{-1} has becomes broaden.

3.2.2 Hydroxyl species and the acidity property

Acidity measurement, type of acid sites and hydroxyl groups in ITQ-6 was characterized by infrared spectroscopy and pyridine molecule was used as probe base.

IR spectrum of delaminated ITQ-6 material evacuated at 400°C shows the formation of large amounts of silanol groups at 3742 cm^{-1} (Figure 3). No band was observed in the pyridine region. After adsorption of pyridine, the silanol band at 3742 cm^{-1} decreased in intensity and four sharp bands appeared assigned to the vibration of physically adsorbed pyridine (1596 cm^{-1}), Brönsted acid site (1545 cm^{-1}), Brönsted and Lewis acids site (1490 cm^{-1}) and also Lewis acid site (1445 cm^{-1}). As the temperature of the thermal treatments increase, more pyridine was desorbed and reappearance of the other hydroxyls occurred.

From Table 2, delaminated ITQ-6 appears to have larger Lewis acid sites than Brönsted acid sites. After desorption of pyridine at 150°C , Brönsted acid sites showed higher acid strength than Lewis acid sites.

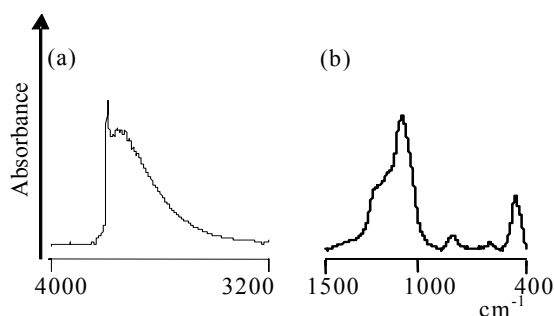


Figure 3 : FTIR spectra of ITQ-6 sample (a) hydroxyl vibration ($P = Ca \approx 10^6\text{ mBar}$) and (b) framework structure.

Table 2 : Acid sites of ITQ-6 measured by Adsorption Desorption of Pyridine at increasing temperatures

Brönsted (μmoleg^{-1})		Lewis (μmoleg^{-1})	
150 °C	400 °C	150 °C	400 °C

ITQ-6 10.14 0.92 18.91 2.60

3.3 Nitrogen adsorption

Figure 4 exhibits nitrogen adsorption-desorption isotherms of ITQ-6 sample. The ITQ-6 exhibits the adsorption-desorption isotherm of Type IV and hysteresis loop of Type H3 which characterized it as mesoporous material. There are two steps involved in the adsorption-desorption isotherm plot. At low relative pressure, the volume of nitrogen adsorbed increased slowly due to the monolayer and multilayer adsorption of nitrogen on the wall. At relative pressure greater than 0.4, the volume of nitrogen adsorbed increased drastically due to the capillary condensation in the mesopores. Hysteresis loop of Type H3 in the range 0.40-1.00 shows that capillary condensation occurred in the slit shape pores with non-uniform size. This proved that the structure in the ITQ-6 material had collapsed.

The pore size distribution in Figure 4 (small box) represents the type of pore exhibits for the ITQ-6 sample. From the plot, ITQ-6 gives a sharp peak in the mesoporous region centered at $C_a = 4$ nm indicates a uniform pore size distribution.

Table 3 summarises the values obtained by applying the BET equation to the values of the nitrogen adsorption isotherm for ITQ-6. The micropore surface area, external surface area and micropore volume were determined from the 't-plot' by using Harkin-Jura equation. From the table it can be proved that the pore with meso size has been formed by delamination process of PREFER to ITQ-6 leads to material with higher external surface. ITQ-6 with higher external area is more accessible for bulk molecule to the acid sites compared to the micropore FER type material.

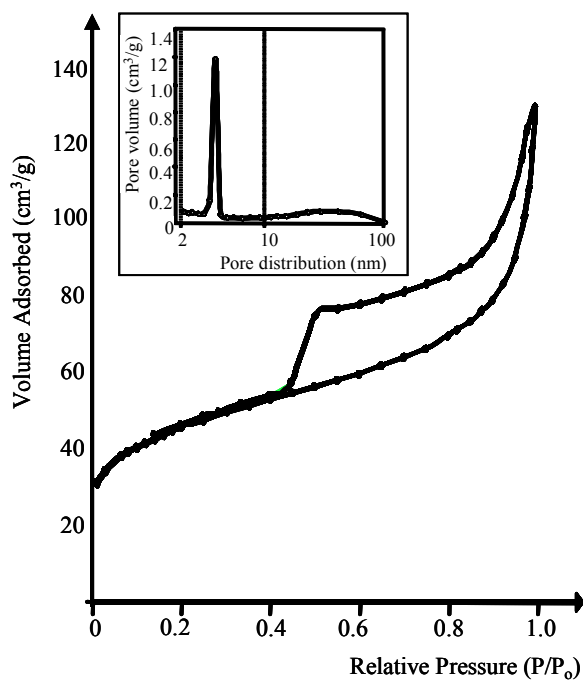


Figure 4 : Nitrogen adsorption-desorption of ITQ-6 (insert shows pore size distribution).

Table 3 : Specific surface areas ($S[m^2g^{-1}]$) and micropore volume ($V[cm^3g^{-1}]$) determined from nitrogen adsorption isotherm.

	S_{Total} (BET)	S_{Micro} (t-plot)	$S_{External}$ (t-plot)	V_{Micro} (t-plot)
ITQ-6	153.94	59.48	94.46	0.025

3.4 Field Emission Scanning Electron Microscope (FESEM)

The morphologies of the PREFER and ITQ-6 samples (delaminated PREFER) are shown in Figure 5.

The crystals of PREFER (Figure 5a) had a plate-like morphology with a very homogeneous distributions. The aggregates are made of very thin stacked crystals (about $60 \times 10 \times 0.5 \mu m$).

The FESEM micrographs of ITQ-6 (Figure 5b) material clearly indicates that the crystal size and morphology of PREFER is strongly influenced by the delamination process. The plate-like crystal was destroyed upon the delamination process. There are no specific structure was observed in ITQ-6 micrographs. The plate-like crystals of FER is still present in the ITQ6 material. This was proved by XRD of ITQ-6 where a sharp peak at $2\theta 9.8^\circ$ indicates the present of small amount of crystal material in ITQ-6 sample.

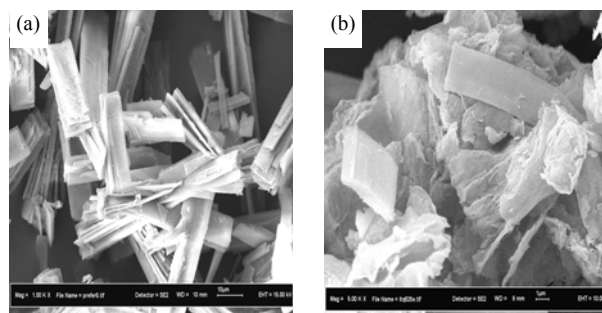


Figure 5 : FESEM micrographs of (a) PREFER and (b) ITQ-6

4. Conclusion

Delamination of layered PREFER material leads to the formation of mesopore material, ITQ-6 which has a laminar character with the lamellae having a microporous internal structure with channels formed by rings of 8 and 10 membered T atoms. Modification of PREFER to the delaminated ITQ-6 can be done by intercalation of bulk molecule between the layers followed by sonication of swollen material and calcinations to remove the organic molecules.

The delaminated ITQ-6 material was found to have higher external surface with lower micropore surface area compared with FER type material. The ITQ-6 also appears to have both Lewis and Brønsted acid sites, which is a characteristic of zeolite acidity.

Higher external surface area and lower micropore surface area of delaminated ITQ-6 with existence of zeolite acidity is thought to be an ideal catalyst for processing bulky molecules.

Acknowledgement

We gratefully acknowledge funding from The MOSTI, under IRPA grant no: 09-02-06-0057 SR0005/09-03 and MARA for the scholarship funding.

References

- [1] Jung, W. Y., Yoo, J. W., Lee, C. W., Jeong, H. C., Park Y. K. and Park S. E. 2000. *t*-Butylation of 1,2-dihydroxybenzene over acidic zeolites. *Catalysis Today* 60: 255–261
- [2] Arishtirova K., Kovacheva, P. and Vassilev, S. 2001. BaO/NaX zeolite as a basic catalyst for oxidativemethylation of toluene with methane. *Applied Catalysis A: General* 213: 197–202
- [3] Javier P. R., Kumar, M. S. and Brückner, A. 2004. Reduction of N₂O with CO over FeMFI zeolites: influence of the preparation method on the iron species and catalytic behaviour. *Journal of Catalysis*, 223: 13-27
- [4] Serrano, D. P., Aguado, J., Escola, J. M. and Rodríguez, J. M. 2005. Influence of nanocrystalline HZSM-5 external surface on the catalytic cracking of polyolefins. *J. Anal. Appl. Pyrolysis* 74: 353–360
- [5] Huang, Y., Meng, X., Dang, Z., Weng, S. and Zhang, C. 1995. Light olefin synthesis from carbon dioxide by hydrogenation over Fe₃(CO)₁₂ supported on ZSM-5 zeolite catalyst. *J. Chem. Soc., Chem. Commun.* 10:1025-1026
- [6] Brendan, C., William M. C. and John, N. 1983. Alkylation reactions over ion-exchanged molecular sieve zeolite catalysts. Part 4. Competitive alkylation of toluene and benzene with ethanol, n-propyl alcohol, isopropyl alcohol and t-butyl alcohol: consideration of the reactivity and selectivity of the alkylating species. *Journal of the Chemical Society, Faraday Transactions 1*, 79: 327 - 342
- [7] A. Corma, Navarro, M. T. and Pariente J. P. 1994. Synthesis of an ultra large pore titanium silicates isomorphous to MCM-41 and its application as a catalyst for selective oxidation of hydrocarbon. *Journal of Chemical Society, Chemical Communication*. 147–148.
- [8] Aguado, J., Serrano, D. P., Escola, J. M. and Rodríguez, J. M. 2004. Low temperature synthesis and properties of ZSM-5 aggregates formed by ultra-small nanocrystals. *Microporous and Mesoporous Materials* 75: 41–49
- [9] Corma, A. Fornes, V., Pergher, S. B., Maesen, Th. L. M. and Buglass, J. G. 1998. Delaminated Zeolite Precursor as Selective Acidic Catalyst. *Nature*. 396: 353-356.
- [10] Climent, M. J., Corma, A. and Velty, A. 2004. Synthesis of hyacinth, vanilla, and blossom orange fragrances: The benefit of using zeolites and delaminated zeolites as catalysts. *Applied Catalysis A: General*. 263: 155-161
- [11] Corma, A., Diaz, U., Domine, M. E. and Fornés, V. 2000. New Aluminosilicate and Titanosilicate Delaminated Materials Active for Acid Catalysis, and Oxidation Reactions Using H₂O₂. *Journal American Chemical Society*. 122(12): 2804 - 2809;
- [12] Corma, A., García, H. and Miralles, J. 2001. High activity layered zeolite ITQ-2 as catalyst for the hydroxyalkylation of 2-methoxynaphthalene and naphthalene with paraformaldehyde; Comparison of its performance with that conventional zeolites or mesoporous Al/MCM-41. *Microporous and Mesoporous Materials*. 43: 161-169.
- [13] Chica, A. Diaz, U., Fornes, V. and Corma, A. 2002. *Delaminated Microporous Solid*. (US 6469226)
- [14] Corma, A., Fornes, V. and Diaz, U. 2001. ITQ-18 a new delaminated stable zeolite. *Chemical Communications*, 2642 – 2643
- [15] Zebib, B., Lambert, J. F., Blanchard, J. and Breyse, M. 2006. LRS-1: A New Delaminated Phyllosilicate Material with High Acidity. *Chem. Mater.* 18: 34-40
- [16] Schreyeck, L., Caullet, P., Mouganel, J. C., Guth, J. L. and Marler, B. 1996. PREFER: A New Layered (Alumino) Silicate Precursor Of FER-Type Zeolite. *Microporous Materials*. 6: 259-271.
- [17] Concepción, P. López, C. Martínez, A. and Puentes, V. F. 2004. Characterization and catalytic properties of cobalt supported on delaminated ITQ-6 and ITQ-2 zeolites for the Fischer–Tropsch synthesis reaction. *Journal of Catalysis*. 228: 321–332

Effect of Temperature on Catalytic Cracking of Palm Oil to Gasoline Over Copper-HZSM-5

Tirena Bahnur Siregar

Faculty of Chemical and Natural Resources Engineering
Universiti Teknologi Malaysia, 81310 UTM, Skudai, Johor, Malaysia
Email: tirenasiregar@yahoo.com

Abstract

The objective of this research was to study the temperature effect on conversion of palm oil to gasoline over HZSM-5, copper loaded HZSM-5. The Cu/HZSM-5 catalyst was prepared using incipient wetness impregnation technique. The crystallinity, form of hydroxyl groups, acidity and surface area were examined by X-ray diffraction (XRD), Fourier transform infrared spectroscopy (FTIR), Pyridine adsorption - infra red spectroscopy (Py-IR) and Nitrogen Adsorption (NA). The results showed that the crystallinity, acidity and surface area of catalyst decreased by loading metals. The catalytic activities were carried out in a fixed bed reactor at atmospheric pressure. Then, the cracking process was conducted in the range temperature between 475°C and 525°C at weight hourly space velocity (WHSV) of 2.5 h⁻¹ for four hours. Loading copper on the HZSM-5 catalyst decreased the catalytic activity. The performance of palm oil cracking over HZSM-5 higher than using Cu/HZSM-5 in terms of conversion and gasoline selectivity at the given range of temperature. The richest component of gasoline was naphthenes followed by aromatics and isoparaffins. Gas yield decreased with copper loading and almost constant with increasing temperature. Gas product consisted mainly C₃ and C₄ compounds, while the production of the gas over Cu/HZSM-5 was higher than HZSM-5.

Keywords: HZSM-5, Cu/HZSM-5, cracking

1. Introduction

The conversion of hydrocarbons to valuable chemicals such as gasoline and aromatic compounds in catalytic processes have been investigated and progressively improved to develop an alternative and renewable source of an environmental friendly liquid fuel. Currently, some researchers are concentrating on developing alternative and renewable sources of an environmental friendly liquid fuel. The medium pore zeolite, ZSM-5 with a pore diameter of 4.5-6.0 Å is a crystalline shape-selective zeolite that has been widely investigated in catalytic conversion of plant oil to fuels and chemicals [1].

Palm oil is enriched with triglycerides that can be converted to clean premium fuels and chemicals. Catalytic cracking using H-ZSM-5, gave high gasoline yield with high aromatic contents [2]. By passing palm oil vapor through H-ZSM-5, high gasoline selectivity enriched with naphthenes and isoparaffins compounds can be obtained [3]. However, the gas yield was high whereas more liquid products are desired.

A modification of the catalyst by transition metal on zeolite is made to enhance the performance of the catalyst. Copper and cobalt-containing catalysts are widely used in redox reaction. Study on performance of Cu-ZSM-5 for selective catalytic reduction (SCR) of NO to N₂ using methane and propane as reduction agent has been conducted. It has been found that the catalytic activity of

both catalysts decreased with increasing metal loading [4]. In addition, Cu(I)-ZSM-5 and Cu(I)-Y converted n-hexane to increase amount of gasoline and aromatics [5]. Furthermore, Tsoncheva *et.al.*, 2004 carried out methane dehydrogenation over Cu-ZSM-5 to produced an alternative fuel or chemicals [6]. They obtained that catalytic activity increase in the present of copper ion.

The objective of this research was to investigate the effect of acidity, surface area, pore size and copper loaded on H-ZSM-5 for palm oil cracking. Effect of the temperature on the catalysts for palm oil to gasoline was also studied.

2. Materials and methods

The study was made on a RBDPO and the chemicals used was nitrogen gas, NH₄-ZSM-5 (Si/Al=30) from Zeolyst International (CBV 3024 E) and the copper (II) nitrate-3-hydrate was employed from Riedel de Haen. HZSM-5 was prepared by drying the ammonium forms of the zeolite overnight and subsequently calcined at 550°C for five hours. Copper was impregnated on the HZSM-5 by the incipient wetness technique and recalcined at 550°C for four hours.

Several characterization techniques were employed in order to determine the characteristics of catalyst used in this study. The XRD analysis was conducted to check zeolite

structure using Siemens Diffractometer D5000 (Software Diffract Plus). The Fourier transform infrared (FTIR) was performed to study a framework of catalyst using Shimadzu 2000 spectrophotometer. The acidity of catalyst was measured by the Pyridine Adsorption Infra Red (Py-IR). The BET surface area, pore size were measured using Micromeritics Accelerated Surface Area and Porosimetry (ASAP) 2010 equipment.

The palm oil cracking was performed in fixed bed reactor under nitrogen gas over the catalyst at atmospheric pressure describe elsewhere [3]. The reaction temperature was in range of 475°C- 525°C. One gram of calcined catalyst powder (particle size about 0.5 mm) was loaded over 0.2 g quartz wool supported by a pin over a stainless steel pipe in a stainless steel reactor (150 mm long and 9 mm ID). The reactor was placed in a vertical furnace and preheated at 450°C for an hour, flushed with N₂ (100 ml/min) for 15 minutes. The reactor was then heated to the desired temperature. Palm oil was injected by a syringe pump with WHSV=2.5 h⁻¹. It was preheated at 120°C and 250°C respectively before entering the reactor. The experiment was conducted for four hours with the N₂ flowing at 1 L/h. The organic liquid product (OLP) was distilled in a micro distillation unit at atmospheric pressure. The gaseous products were analyzed with a Hewlett Packard Agilent 6890N GC system equipped with a thermal conductivity detector (TCD). The gasoline product was analyzed with a gas chromatograph (Perkin Elmer, Auto System) using capillary column (100m long x 0.25mm width, model no: SGE 100QC2.5 BPI PONA) with a flame ionization detector (FID). A substantial amount of coke was accumulated on the catalyst after each experiment. The spent catalyst was weighed using a microbalance and heated to 500°C in a furnace for five hours. The difference in the weight of the spent catalyst before and after heating was termed as the weight of coke, which was burned off.

The activity of catalytic cracking calculation based on feedstock. The conversion, gasoline selectivity and yield are defined as follow:

$$\text{Conversion (wt\%)} = \frac{F - R}{F} \times 100\% \quad (1)$$

$$\text{Selectivity (wt\%)} = \frac{P}{F - R} \times 100\% \quad (2)$$

$$\text{Yield (wt\%)} = \frac{P}{F} \times 100\% \quad (3)$$

where F is the palm oil feed weight (g), R is the residual oil weight (g) and P is the product weight (gasoline, heavy product, residue product, gas and coke).

3. Result and discussion

The XRD patterns of Cu/HZSM-5 at various percentages of copper loading from 2 wt% to 6 wt% are

demonstrated in Figure 1. The peaks at 2θ were of 35.5° and 38.7° correspond to the (002) and (111) planes of divalent copper oxide (CuO) (JCPDS Pattern No. 2-1040). It was observed that CuO peak appeared clearly for 6 wt%.

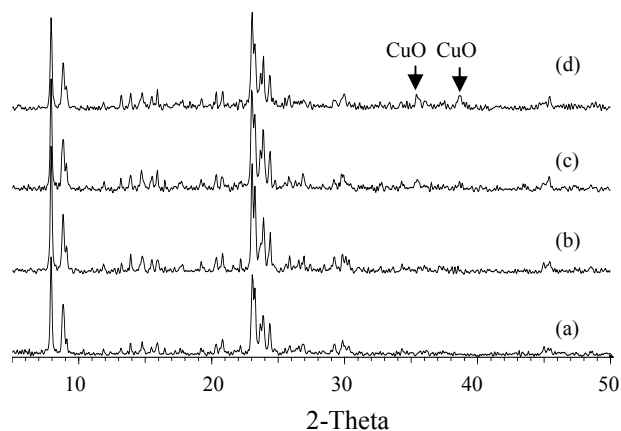


Figure 1. XRD pattern of (a) H-ZSM-5, (b) 2 wt% Cu-ZSM-5, (c) 4 wt% Cu-ZSM-5 and (d) 6 wt% Cu-ZSM-5

The FTIR spectrum of HZSM-5 and Cu/HZSM-5 is shown in Figure 2. The structure sensitive absorption around 1200 cm⁻¹ and 550 cm⁻¹ was a special interest to distinguish the zeolite types. The frequency band at 1099 cm⁻¹ was assigned to asymmetric stretching of framework Si-O-Si or Si-O-Al bonds. Band at 796 cm⁻¹ is due to presence of non-bonded oxygen Si-O⁻, while the bands of 547 cm⁻¹ and 451 cm⁻¹ also corresponded to bending vibrations of O-Si-O bonds [7]. The spectrum of HZSM-5 was similar to the spectrum of Cu/HZSM-5, whereas the vibration of CuO bonds that appeared at 575 cm⁻¹, 500 cm⁻¹ and 460 cm⁻¹ could not be observed due to the presence of a broad band of 547 cm⁻¹ from HZSM-5 [8].

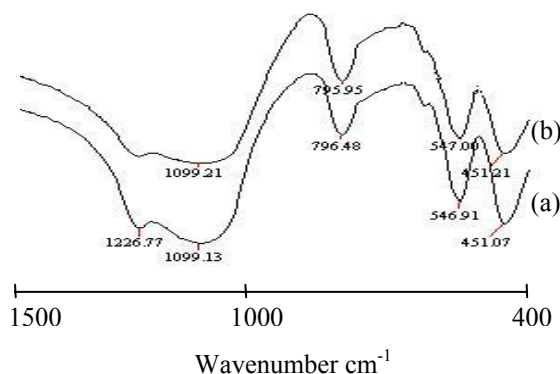


Figure 2. FTIR spectrum of (a) HZSM-5 and (b) Cu/HZSM-5

The physicochemical properties of HZSM-5 and Cu/HZSM-5 are tabulated in Table 1. The crystallinities of HZSM-5 and 6 wt% Cu/HZSM-5 were 100% and 96% respectively. The slight difference of the crystallinity might be due the instrumentation error. The physicochemical property of Cu/HZSM-5 was different compared to HZSM-

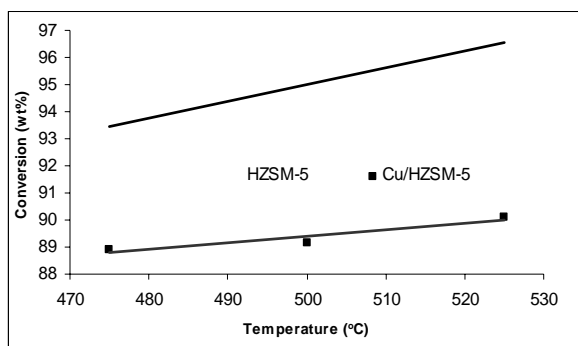
5 in terms of their Brönsted and Lewis acidity. The Brönsted acidity of Cu/HZSM-5 decreased whereas the Lewis acidity increased. This is probably due to increased amount of non-framework aluminium or hydrogen cation that was replaced by the metal ion (Cu^{2+}). The decrease in H^+ ions decreases the Brönsted acidity, meanwhile the Lewis acidity was created by Cu^{2+} ions [9]. The surface area of HZSM-5 was larger than Cu/HZSM-5 and pore size of Cu/HZSM-5 was bigger than HZSM-5. The decrement in the surface area indicated that CuO covered the surface of HZSM-5 zeolite. The slightly increment in pore size of Cu/HZSM-5 might be due to attach of copper species in the micropores of HZSM-5 [10].

Table 1: Physicochemical properties of HZSM-5 and Cu/HZSM-5.

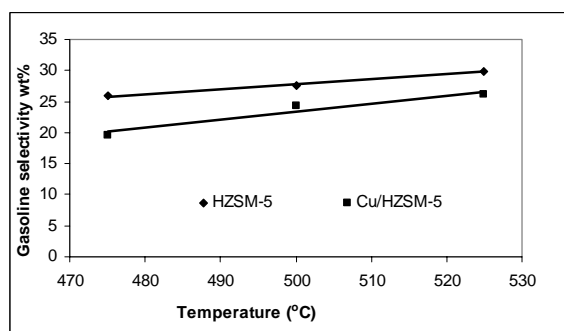
Catalyst	Property				
	C	BSA	BPS	BA	LA
HZSM-5	100	333.32	25.11	134	61
Cu/HZSM-5	96	295.83	26.58	68	363

where **C** is cristallinity, **BSA** is BET Surface Area (m^2/g), **BPS** is BET Pore Size (Å), **BA** is Bronsted Acidity ($\mu\text{mol}/\text{g}$), and **LA** is Lewis Acidity ($\mu\text{mol}/\text{g}$)

The catalytic cracking activities of palm oil over HZSM-5 and Cu/HZSM-5 from 475 o C to 525 o C at WHSV of 2.5 h-1 are demonstrated in Figure 3 to 5.



(a)

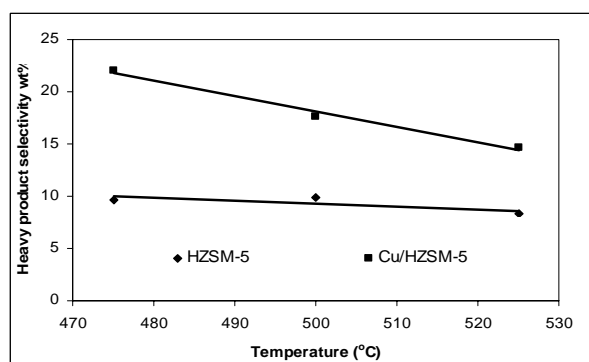


(b)

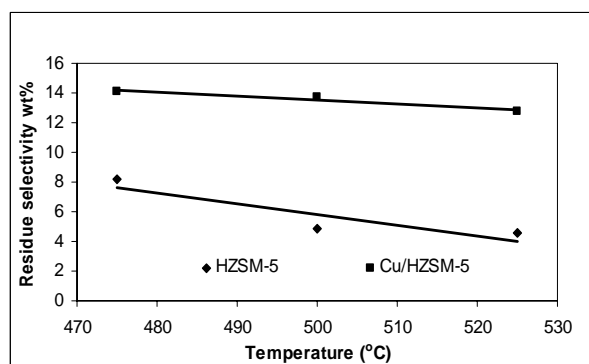
Figure 3. The catalytic activity of palm oil cracking over HZSM-5 and Cu/HZSM-5 at various temperatures and WHSV = 2.5 h⁻¹: (a) conversion and (b) gasoline selectivity.

The conversion and gasoline selectivity of palm oil cracking increased with increasing temperature, which at high temperature of 525 °C using HZSM-5 were 96.12 wt% and 29.92 wt%, respectively. Furthermore, the conversion and gasoline selectivity at the same operation condition obtained over Cu/HZSM-5 were lower than the parent catalyst HZSM-5. Using Cu/HZSM-5 at high temperature of 525°C obtained the conversion and gasoline selectivity were 90.12 wt% and 26.15 wt%, respectively. The surface area of the parent HZSM-5 was larger than copper loaded HZSM-5. Hence the conversion over HZSM-5 was higher than Cu/HZSM-5. In addition, the Brönsted acidity also plays an important role in hydrocarbon cracking [11]. Consequently, the cracking activity in term of conversion and gasoline selectivity of Cu/HZSM-5 was reduced due to lower concentration of Brönsted acid sites.

Figure 4a and 4b show the heavy product selectivity and residue selectivity of the palm oil cracking over HZSM-5 and Cu/HZSM-5.



(a)



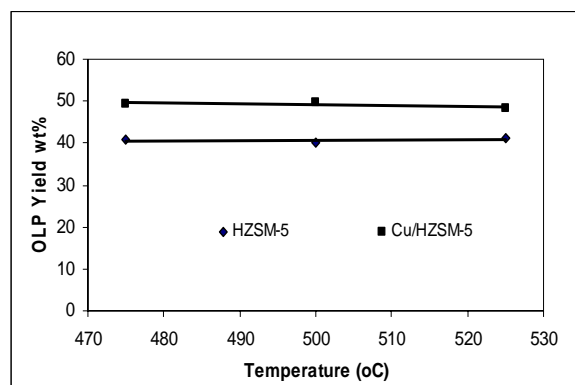
(b)

Figure 4. The palm oil cracking over HZSM-5 and Cu/HZSM-5 at various temperatures and WHSV = 2.5 h⁻¹: (a) heavy product selectivity and (b) residue selectivity.

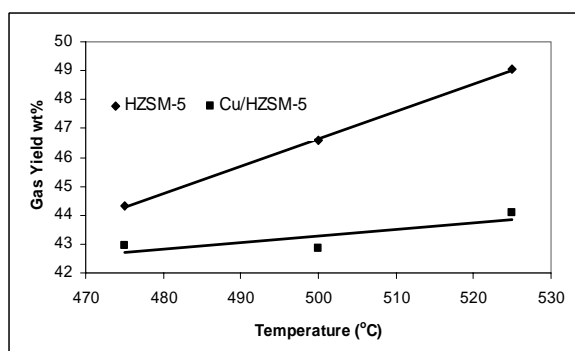
Selectivity towards heavy product over HZSM-5 and Cu/HZSM-5 declined with increasing temperature. The heavy product selectivity at 525°C were 8.3 wt% and 14.62

wt%, respectively. The gasoline selectivity was contrary to the heavy product selectivity [12]. Increase in secondary reaction such as cyclization and isomerization occurred at high temperature, thus contributing to the high amount of gasoline. Twaiq *et al.*, 1999 obtained that at WHSV of 2.5 h⁻¹ the heavy product (kerosene + diesel) decreased with increasing temperature over HZSM-5 [13]. They gained at 450°C was 15.5 wt%, while in this study at 475°C and WHSV of 2.5 h⁻¹ over HZSM-5 was about 9.64 wt%. Selectivity towards residue was directly the opposite of the conversion. The conversion increased with increasing temperature, whereas the selectivity towards residue decreased with increasing temperature. Then metal loading uncompleted cracking process with producing of high residual oil. The residue selectivity at 525 °C were 4.54 wt% and 12.75 wt%, respectively.

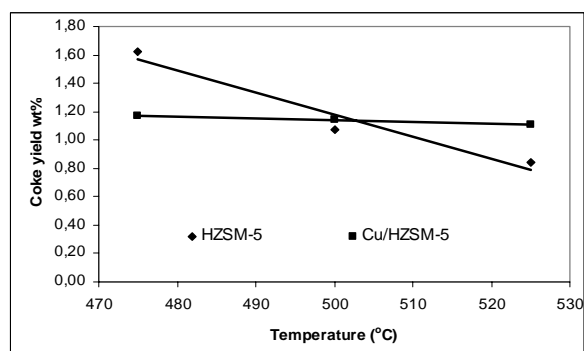
Figure 5a, 5b and 5c show the OLP, gas and coke yield of the palm oil cracking over HZSM-5 and Cu/HZSM-5. The yield of product included OLP, gas and coke that produced by cracking of palm oil at various temperature over HZSM-5 and Cu/HZSM-5 catalysts. The OLP yield was almost constant with increasing temperature, higher with metal loading. Yield of gas increased with increasing temperature, while coke yield decreased. Increasing of gas yield at high temperature could be due to increased in secondary cracking reaction. Partial gasification of coke could also be a factor to the huge gas fraction and low coke fraction [12]. Gas yield decreased with copper loading and increased with increasing temperature. Gas product consisted mainly of C₃ and C₄ compounds, while HZSM-5 produced higher than Cu/HZSM-5.



(a)



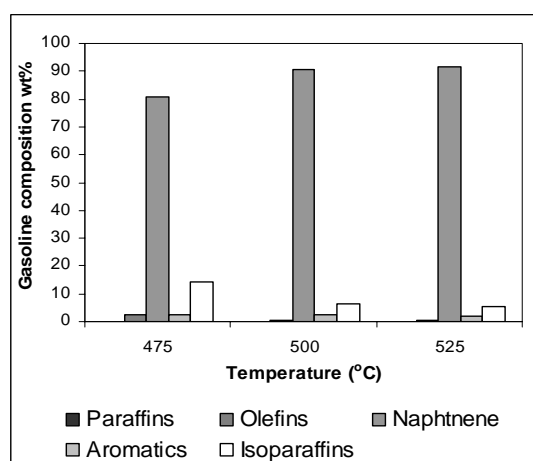
(b)



(c)

Figure 5. The yield product of palm oil cracking over HZSM-5 and Cu/HZSM-5 at various temperatures and WHSV = 2.5 h⁻¹: (a) OLP yield, (b) gas yield and (c) coke yield.

The gasoline composition at various temperatures over HZSM-5 and Cu/HZSM-5 are figured in Figure 6a and 6b. The gasoline enriched mainly of naphthenes, beside aromatics and isoparaffins. The naphthenes was consisted mostly in gasoline that produced by cracking palm oil over HZSM-5. This component increased with increasing temperature. The high naphthenes' gasoline composition in this study might be due to the dehydrogenation of paraffins produced olefins by pyrolysis of preheated palm oil at 250°C before it entering the reactor. **Then through oligomerization and cyclization produced naphthenes.** The copper loading decreased the naphthenes and increased the aromatics. The increasing of aromatics compound might be due to increasing the Lewis acidity. The naphthenes were yielded mainly of 1t, 2c, 3-trimethylcyclohexane, while the aromatics were found to consist of alkylbenzene only.



(a)

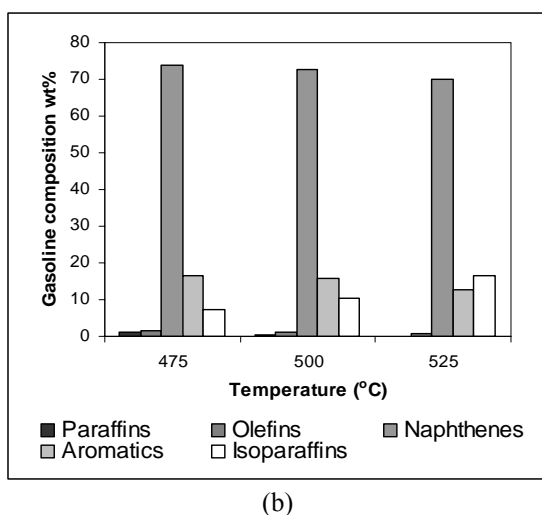


Figure 6. The gasoline composition of palm oil cracking over (a) HZSM-5 and (b) Cu/HZSM-5.

Twaiq, 2004 carried out the palm oil conversion to liquid fuel over HZSM-5 and Ni/HZSM-5 at 450°C and WHSV of 2.5 h⁻¹ with oil/cat. ratio of 7.2 wt% and particle size of 32 µm, which the catalyst prepared by ion exchanged technique. The parent catalyst HZSM-5 has Si/Al ratio of 50 [14]. The surface area of HZSM-5 and Ni/HZSM-5 were 400 m²/g and 435 m²/g, respectively. They obtained that conversion of palm oil over HZSM-5 and Ni/HZSM-5 were 96.90 wt% and 99.16 wt%, respectively, while gasoline yield were 44.62 wt% and 41.51 wt%. Furthermore, the aromatics compound increased with metal loading, whereas gasoline enriched of BTX aromatics. In addition the gas yield gained about 17.46 wt% and 27.47 wt%, respectively. Although the conversion increased with metal loading of Ni, nevertheless it effected the decreasing in gasoline yield. The conversion and gasoline selectivity of their results were higher than this study, it might be due to higher Brønsted acidity and larger surface area [11]. In addition the heavy product (kerosene + diesel) yield and the gas yields gained were and 17.4 wt%, respectively. However, in this study the C₁ of the gas product was found to be lower of the Twaiq findings, but CO₂ was found to be higher and naphthenes enriched in gasoline composition.

4. Conclusion

Catalytic cracking of palm oil over Cu/H-ZSM-5 catalysts showed that catalyst activity depends on the Brønsted acid sites and surface area. Therefore, HZSM-5 without metal loading gave better performance than that of metal loading at any temperature in terms of conversion and gasoline selectivity. Conversion and gasoline selectivity increased with increasing temperature and declined with metal loading. The highest conversion and gasoline selectivity obtained at 525°C were 96.12 wt% and 29.92 wt%, respectively by using HZSM-5. Using Cu/HZSM-5 the palm oil conversion decreased to 90.12 wt% while gasoline selectivity also decreased to 26.15 wt%. Therefore,

HZSM-5 alone was suitable for palm oil cracking to gasoline. The consisted naphthenes in gasoline decreased while the aromatics increased due to the increasing of acidity. The most significant component in the gasoline was naphthenes followed by aromatics and isoparaffins. Gas yield decreased with copper loading and increased with increasing temperature. Gas product consisted mainly of C₃ and C₄ compounds while HZSM-5 produced higher than Cu/HZSM-5.

References

- [1] Haag, W. O., Rodewald, P. G. and Weisz, P.B. 1980. Catalytic Production of Aromatics and Olefins from Plant Materials. *Symposium of Alternate Feedstocks for Petrochemicals*. San Francisco. 650-656.
- [2] Tan, Y. L., Abdul Rahman Mohamed and Bhatia, S. 1999. Catalytic Conversion of Palm Oil to Fuels and Chemicals. *Canadian Journal of Chemical Engineering*. 77 : 156-162.
- [3] Siregar, T. B., Amin, N. A. S. and Karim, S. K. 2005. Catalytic Conversion of Palm Oil to Gasoline over HZSM-5: Effect of Temperature and Hydrotreatment of catalysts. In *Proceedings of Annual Fundamental Science Seminar 2005*. Universiti Teknologi Malaysia. Johor.
- [4] Seyedeyn-Azad, F. and Zhang D. 2001. Selective Catalytic Reduction of Nitric Oxide over Cu and Co ion-exchange ZSM-5: Effect of SiO₂/Al₂O₃ Ratio and Cation Loading. *Catalysis Today*. 68: 161-171.
- [5] Amin, N. A. S., Tan, E. F., Law, S. W. and Wong, C. T. 2002. *Proceeding of the 2nd World Engineering Congress*. Kucing, Malaysia. 237-243.
- [6] Tsoncheva, T., Venkov, Tz., Dimitro, M., Minchev, C. and Hadjivanov, K. 2004. Copper Modified Mesoporous MCM-41 Silica: FTIR and Catalytic Study. *Journal of Molecular Catalysis A: Chemical*. 209: 125-134.
- [7] Le Van Mao, R., Le, T. S., Fairbain, M., Muntasar, A., Xiao, S. and Denes, G. 1999. ZSM-5 Zeolite with Enhanced Acidic Properties. *Applied Catalysis A: General*. 185: 41-52.
- [8] Gervasini, A. 1999. Characterization of The Textural Properties of Metal Loaded ZSM-5 Zeolite. *Applied Catalysis A: General*. 180 : 71-82.
- [10] Nicolaides, C. P. 1999. A Novel Family of Solid Acid Catalysts: Substantially Amorphous or Partially Crystalline Zeolitic Material. *Applied Catalyst A: General*. 185 : 211-217.
- [11] Corma, A., Fornes, V., Mocholi, F.A., Monton, J.B. and Rey, F. 1991. Influence of Superacid Sites in Ultrastable Y Zeolites on Gas Oil Cracking. In: Ocelli, M.L. (Ed.) *Fluid Catalytic Cracking II: Concepts in Catalyst Design*. Washington, D.C : American Chemical Society. 12-25.
- [12] Adjaye, J.D. and Bakhshi, N.N. (1995). Production of Hydrocarbons by Catalytic Upgrading of a Fast Pyrolysis Bio-Oil. Part I: Conversion Over Various Catalysts. *Fuel Processing Technology*. 45: 161-183.

- [13] Twaiq, F. A., Zabidi, N. A. M. and Bathia, S. (1999). Catalytic Conversion of Palm Oil to Hydrocarbons: Performance of Various Zeolite Catalysts. *Ind. Eng. Chem. Res.* 38: 3230-3237.
- [14] Twaiq, F. A., Zabidi, N. A. M., Mohamed A. R. and Bathia, S. 2004. Performance of Composite Catalyst in Palm Oil Cracking for the Production of Liquid Fuels and Chemicals. *Fuel Processing Technology.* 85: 1283-1300.

Characterization of Acid-Treated Carbon Nanotubes (CNTs) Synthesized Using MgO and CaCO₃ Supports

Mohamed Nuruddin bin Mohamed Nasir¹, Nor Aziah bte Buang*¹,

¹Faculty of Science, Universiti Teknologi Malaysia,
81310 UTM Skudai, Johor, Malaysia.
Tel. No. (Office): 07-5533333 (ext: 34123) Fax: 07-5566162
E-mail: noraziah@kimia.fs.utm.my

Abstract

Carbon nanotubes have been synthesized using catalytic chemical vapor deposition (CCVD) method. Magnesium oxide and calcium carbonate powders coated with a mixture of ferum acetate and cobalt acetate salts were put inside a CCVD reactor in order to become the support for the CNT growth. The carbon source used was C₂H₂ flowed into the reactor at a specified rate to stimulate CNT synthesis. The nanotubes are then treated with nitric acid. The as produced and treated CNTs were analyzed with scanning electron microscope (SEM), field-emission SEM (FE-SEM), energy dispersive X-ray analysis (EDAX) single-point BET analysis and thermogravimetric analysis in order to compare them. The analysis results show that acid treatment is capable of removing the catalyst particles and catalyst support from the magnesium oxide and calcium carbonate-supported CNTs. The surface area of the CNTs samples were also found to be increased after the purification process.

Keywords:

Carbon nanotubes (CNTs), CCVD, alkaline-earth catalyst supports, Fe/Co catalysts

1. Introduction

Since its discovery and successful synthesis by Iijima in 1991 [1], CNTs has been the focus of a virtual storm of research, both to better understand its unique properties and to harness its potential in commercial applications such as hydrogen storage, atomic force microscope probe, microelectronic transistor, electrical field emitter of flat panel display and scanning tunneling microscope tip have been stimulated tremendously. High-quality and well-aligned carbon nanotubes are essential to the potential applications in the field of microelectronic industries. CNTs are unique tubular structures with nanometer diameters and large length/diameter ratios (~ 1000). CNTs can be considered conceptually as a prototype one-dimensional (1D) quantum wire [2]. The fundamental building block of carbon nanotubes is the very long all-carbon cylindrical single-walled carbon nanotubes (SWNTs), one atom in wall thickness and tens of atoms around the circumference (typical diameter ~ 1.4 nm). A single-walled carbon nanotube can be described as a graphene sheet rolled into a seamless hollow cylindrical shape with a high degree of molecular perfection. A multi-walled carbon nanotube consists of concentric cylinders with an interlayer spacing of 34 nm and a diameter of typically 10 – 20 nm.

There are several methods of synthesizing CNTs, namely arc-discharge, laser ablation, and chemical vapor deposition (CVD). One of the most popular variations of the CVD method is the chemical CVD (CCVD) due to its

relative simplicity and high yield. CCVD synthesis is carried out by pyrolysis of hydrocarbons using metal catalysts supported on inert materials with high surface area such as zeolites and aluminum oxide. This is potentially the most practical and cost-effective way of synthesizing CNTs on a large scale. However, this approach has a few drawbacks, namely the purity of the CNTs is quite low and the supports are quite difficult to remove due to its chemically stable nature. Therefore, suitable methods for purification of the as-synthesized CNTs have been researched [3-5]. This includes attempts to synthesize CNTs using supports that are more easily removed yet still retain a high surface area, such as Ca and Mg salts [6]. The present work involves the application of MgO and CaCO₃ salts as supports in a CCVD synthesis of CNTs with iron–cobalt catalysts, and the effectiveness of a simple acid-wash purification of its resulting CNTs. The effects of acid washing on the surface area of CNTs will also be investigated in detail.

2. Experimental Details

There are several steps involved in obtaining the desired CNT samples and analyzing them. The first step is the preparation of supported catalysts that will be used in the CCVD system. The second step is the synthesis of CNTs using the prepared catalysts in the custom-built CCVD system. The third step is purification of the CNT samples by acid washing and the last step is analyzing all

the samples obtained (as-produced samples, purified samples).

2.1 Preparation of supported catalysts

The MgO and CaCO₃ supports were dried in an oven at a temperature 110 °C for 24 hours. The acetate salts of Fe and Co were measured and used in the amount needed to achieve the desired weight ratio between the metals and the support and then dissolved in distilled water. When the support was mixed with the salts solution, a paste that absorbed the solution completely was formed. This paste was then dried in an oven at a temperature of 110 °C for 24 hours to remove all the water molecules. The resulting material was a dry, crusty paste which was ground into powder with a mortar and pestle. Finally, the dried catalyst powder was put into waterproof and airtight sample bottle.

2.2 Synthesis of CNTs

In this work, carbon nanotubes were synthesized in a custom-built CCVD system, which incorporates a horizontal tube reactor furnace system. The supported catalysts were placed in a ceramic boat located in the middle of a reactor tube. The tube was then placed in the tube furnace in such a way that the boat containing the catalysts would be in the center of the heating element tube. Acetylene (C₂H₂) and nitrogen (N₂) were introduced into the tube reactor at a pre-determined flow rate. The N₂ acts as a carrier gas for the carbon source gas C₂H₂ and provides the inert condition for the system. The reaction was initiated at ambient pressure and temperature. The carrier gas was flowed into the system for about 15 minutes prior to heating to ensure that air was completely purged from the system. Then, the furnace was then heated up gradually until it reached the preset reaction temperature of 700 °C. Once this temperature was reached, the carbon source was flowed into the system to initiate the carbon deposition process. The reaction was continued for 30 minutes and then was cooled back gradually to room temperature. The carbon source gas flow was also stopped after 30 minutes. Once the system had reached ambient temperature, the carrier gas flow was stopped and the ceramic boat containing the carbon deposit was taken out of the reactor tube for analysis.

2.3 Acid washing of CNTs

In order to determine the viability of MgO and CaCO₃ as supports in the production of CNTs that can be purified with acids, a simple acid washing was done to the CNT samples. The samples were immersed completely in a 5M nitric acid solution for 2 hours. Then, the samples were filtered out of the solution and dried for 24 hours. After the samples were completely dried, they were stored in a waterproof, airtight sample bottle.

2.4 Characterization of CNTs

Characterization of the samples obtained is important to ensure that meaningful data can be extracted from the experiments. In this work, the important information to be obtained is the effectiveness of MgO and CaCO₃ as supports in CNT synthesis and the effectiveness of nitric acid in removing the metal and supports, and also the effects of purification on the surface area of the CNT samples. Therefore, the characterization techniques chosen for this work are SEM (for confirming the existence of CNTs in the carbon deposits), EDAX (for analyzing the elemental composition of the samples), FE-SEM (for the analysis of CNT morphology in greater magnification), and BET surface area analysis (for detecting changes in the surface area).

From SEM and FE-SEM micrographs, the tubes and catalyst particles can be seen clearly at the appropriate magnification. Any surface defects can also be detected using this analysis. Since the EDAX analysis is done with the same instrument as the SEM microscope, it can be done at the same time as the SEM analysis. To initiate the EDAX analysis, different software was used on the computer interface. The software used for EDAX on SEM is the EDAX Control and for the FE-SEM is the INCA Material Analysis. The data obtained is the elemental species present on the surface of the sample, and presented in graph and table forms. For BET analysis, a small amount of the sample (from 0.02 to 0.05 grams) is analysis with the Micromeritics PulseChemiSorp 2705 instrument. The data for as-produced and treated samples are then tabulated for comparison.

3. Results and Discussions

The main objectives of this work are to determine whether the CNT synthesis is successful using the selected supports (CaCO₃ and MgO), to gauge the effectiveness of the acid washing to remove the metal catalyst and supports, and to see the effects of acid washing on the sample. SEM and EDAX analysis will be useful in fulfilling the first and second objectives while BET analysis will help in achieving the third.

3.1 SEM analysis

Figure 1 show the SEM micrographs of samples after acid washing. These micrographs indicate that there are indeed CNTs deposited on the surface. The CNTs produced with Fe/Co-MgO catalyst (Figure 1a) shows CNTs produced are of similar lengths, diameter and density, with bundling occurring in several locations. The bundles are believed to occur because of the van der Waals interaction between the walls of the nanotubes [7]. The growth of CNTs on the support was prolific, with large bundles of CNTs growing out of support granules and smaller group of CNTs growing from other surfaces of the support. The diameters of the CNTs are also quite small, ranging between 15 to 20 nm. Upon closer inspection, the outer walls of the nanotubes are found to be covered with

bumps emanating from inside the tubes, as indicated by the arrows. This is caused by the carbonaceous matter that forms as a by-product of acid treatment [8]. The CNTs from Fe/Co-CaCO₃ catalyst, on the other hand, shows long and uniformly sized tubes in great density, covering most of the surface. The micrograph of the sample in Figure 1b reveals that the same bumps that are observed in the other treated CNTs are also evident here, but in fewer quantities than others. This indicates a higher purity of CNTs. With more refining of the purification technique, a highly pure CNT sample can be isolated. The diameter of the tubes is estimated to be around 20 to 30 nm. These results show that MgO and CaCO₃ are suitable supports for CNT synthesis.

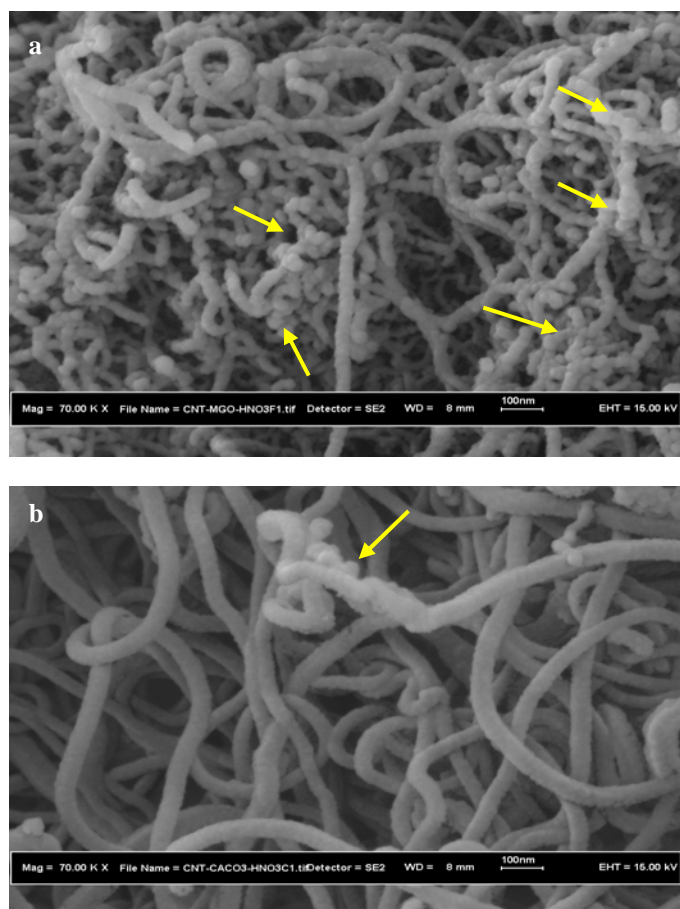


Figure 1: FE-SEM micrograph of CNTs produced by CCVD on Fe/Co supported on: (a) MgO; (b) CaCO₃. Arrows indicate bumps on CNTs

3.2 EDAX analysis

Table 1 shows the elemental composition of CNTs for all supports before and after the acid treatment, which are obtained via EDAX. As can be seen from the table, the treatment with acid managed to remove almost all of the Fe and Co particles in the CNTs. Moreover, the acid treatment was able to remove almost all of the CaCO₃ and MgO supports from the CNTs. This proves that a significant amount of impurities (in the form of metal

catalysts and MgO/CaCO₃ supports) can be removed easily with a simple acid wash method.

Table 1: Elemental composition of CNTs before and after acid treatment

Sample	Elemental composition (wt %)								
	C	Si	Al	Mg	Ca	O	Fe	Co	Others*
Fe/Co-MgO CNTs BEFORE acid wash	44.05	-	-	24.52	-	18.72	6.65	6.05	-
Fe/Co-MgO CNTs AFTER acid wash	66.69	-	-	7.06	-	26.25	-	-	-
Fe/Co-CaCO ₃ CNTs BEFORE acid wash	27.56	-	-	-	40.97	29.21	0.51	1.75	-
Fe/Co-CaCO ₃ CNTs AFTER acid wash	88.93	-	-	-	10.26	0.81	-	-	-

3.3 BET analysis

Table 2 shows the surface area of the samples obtained from the BET analysis. As can be seen from the table, there is a significant increase in surface area of the samples after the acid washing process. This is because most of the catalyst particles, supports and carbonaceous matter had been removed by the HNO₃, which acts not only as a strong acid, but also as an oxidizing agent [9]. The removal of support matter causes more CNTs to be freed from the bundles, increasing the surface area. The increase is particularly drastic in CNTs with CaCO₃ as supports, with a surface area increase of 171.04 percent. CNTs with MgO supports also show a significant increase in surface area with 60.62 percent.

Table 2: Surface areas of CNTs before and after acid wash

Sample	Surface Area (m ² /g)	
	Before	After
Fe/Co-MgO CNTs	75.11	120.64
Fe/Co-CaCO ₃ CNTs	58.23	157.83

Conclusions

This work has shown the effects of acid washing on the physical properties of carbon nanotubes synthesized using the CCVD method with alkaline metal salts as supports. The CNTs produced were of good quality, with uniformly small diameter and smooth walls. A simple acid washing proves to be adequate in removing supports, catalyst particles and carbonaceous matter, vastly

improving the purity of the CNTs and increasing the surface area. A more precise acid treatment method should increase the quality of the CNTs even further.

Acknowledgements

The authors would like to thank Professor Ahmad Fauzi Ismail and colleagues in the Membrane Research Unit (MRU), UTM for their support in giving guidance and allowing the use of their facilities during the course of this work, and MOSTI for funding this project under the IRPA mechanism (Vot 74525) of the National Fuel Cell project.

References

- [1] S. Iijima (1991), *Nature* 354: 56
- [2] Dresselhaus, M.S., Dresselhaus, G. and Avouris, P. (2001); *Carbon Nanotubes: Synthesis, Structure, Properties and Application*; Germany, Springer-Verlag, Chap. 3, pp. 21
- [3] Ko, F.H., Lee, C.Y., Ko, C.J., Chu, T.C., (2004), *Carbon* 43: 2345-2350
- [4] Sato, Y., Ogawa, T., Motomiya, K., Shinoda, K., Jeyadevan, B., Tohji, K., (2001), *J. Phys. Chem. B* 105: 3387-3392
- [5] Huang W., Wang Y., Luo, G., Wei, F., (2003), *Carbon* 41: 2585-2590
- [6] Kathyayini, H., Nagaraju, N., Fonseca, A., Nagy, J.B., (2004), *J. Molec. Cat. A Chemical* 223: 129-136
- [7] Dai, H., Kong, J., Zhou, C., Franklin, N., Tomblor, T., Cassell A., Fan, S. And Chapline, M. (1999), *J. Phys. Chem. B* 103: 11246-11255.
- [8] Chen, P., Zhang H.B., Lin G.D., Hong Q., Tsai K.R., (1997), *Carbon* 35: 1495-1501.
- [9] Li, W., Bai, Y., Zhang, Y., Sun, M., Cheng, R., Xu, X., Chen, Y., and Mo, Y. (2005), *Synthetic Metals*. 155: 509-515.

Synthesis of Highly Reactive Electrogenerated Zinc for Oxidative Addition of Organic Halides

Murni, S.¹, Aishah, A. J.¹, Sugeng, T.², Muhd. Nazlan, M. M.³ and Masao, T.⁴

¹Faculty of Chemical and Natural Resources Engineering,

²Ibnu Sina Institute for Fundamental Science Studies,

³Department of Chemistry, Faculty of Science,

Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Bahru, Johor.

Tel: +60-7-5535523, Fax: +60-7-5581463, Email: aishah@fkkksa.utm.my

⁴Division of Molecular Chemistry, Graduate School of Engineering,

Hokkaido University, Sapporo 060-8628, Japan.

Abstract

Highly reactive electrogenerated zinc has been synthesized by electrolysis of a DMF solution containing naphthalene as a mediator in a one-compartment cell fitted with a platinum cathode and a zinc anode at 0°C. This reactive zinc (EGZn/Naph) could transform an organic bromide to its corresponding organozinc compound in high yield which can not be achieved by the use of either commercial zinc or even usual electrogenerated zinc (EGZn) metal. However, 2 equivalent of naphthalene with respect to EGZn is required to provide sufficient reactivity of zinc in order to convert alkyl bromide to its corresponding organozinc compound. Recently we found that only 0.1 equivalent amount of phenanthrene, anthracene or pyrene with respect to EGZn is more enough to convert ethyl-4-bromobutyrate to its corresponding organozinc compound in 87-95% of yield. Cross-coupling reaction of the organozinc compounds that prepared from EGZn/Naph and alkyl bromides, with aryl iodides in the presence of 5 mol% of palladium catalyst gave the coupling product in high yields.

Keywords:

Electrolysis, reactive zinc, organic bromide, cross-coupling, palladium catalyst

1. Introduction

Recently, research on organozinc compound become more significant because of its compatibility towards wide spectrum of functional group on the organic and has high tolerant to many functional group such as ester, nitrile, etc (Knochel et al., 1992). Organozinc halides can usually be prepared by direct insertion of zinc metal into organic halides (Knochel et al., 1999), but commercially available zinc metal is generally poorly reactive. Therefore, activation of the metal is necessary for the preparation of organozinc halides. Various methods of zinc activation, such as the reduction of zinc halide with alkaline metal or alkali metal naphthalenide, have been reported (Zhu and Rieke, 1991; Hanson and Rieke, 1995). These methods, however, require high temperature and long reaction times, or vigorous stirring during the reaction.

Previously, Tokuda et al. (1993) reported a new method for the preparation of reactive zinc by electrolysis of a DMF solution containing 0.1M Et₄NClO₄ with a platinum cathode and a zinc anode (Scheme 1). It was shown that this electrogenerated reactive zinc (EGZn) was an aggregation of very fine crystalline zinc particles with a

large surface area (Tokuda et al., 1996). It was very reactive and was successfully used in isoprenylation (Tokuda et al., 1993) and allylation (Tokuda et al., 1995; 1996) of aldehydes and ketones. We have also reported a facile preparation of organic compounds from functionalized alkyl iodides by using EGZn and their cross-couplings with aryl halides (Scheme 2) (Kurono et al., 1999).

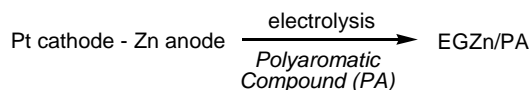
Recently, we developed a new electrochemical method for the preparation of more highly reactive zinc (EGZn/Naph) by using naphthalene as a mediator (Scheme 1) (Aishah, et al., 2001). This more highly reactive zinc successfully transformed bromoalkanes into the corresponding organozinc bromides, which can not be achieved by the use of usual zinc metals, or even if the reactive EGZn was used. The subsequent cross-coupling of corresponding organozinc bromides with various aryl halides in the presence of palladium catalyst gave the corresponding cross-coupling products in high yields (unpublished results). However, in preparation of EGZn/Naph 2 equivalents of naphthalene were required to acquire a sufficient reactivity of EGZn. We examined various types of polyaromatic compounds that can be used as a mediator instead of naphthalene. Therefore, in this paper the authors report a facile preparation of EGZn/PA by

using various types of polyaromatic compound as a mediator in order to convert ethyl 4-bromobutyrate to its corresponding organozinc compound.

2. Methodology

2.1 Preparation of electrogenerated zinc

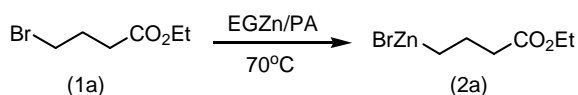
A normal one-compartment cell equipped with a magnetic stirrer bar and serum cap was used. 0.1M tetraethylammonium perchlorate and polyaromatic compound (PA) as a mediator are dissolved in an *N,N*-dimethylformamide solution in the compartment cell fitted with a platinum plate cathode (2×2 cm²) and a zinc plate anode (2×2 cm²). Electrolysis was carried out at 0°C at a constant current of 60 mA/cm² under nitrogen atmosphere. Electrolysis will be continued until 2 F/mol of current passing through the electrolyte (Scheme 1).



Scheme 1

2.2 Preparation of Organozinc Halides

A solution containing EGZn/PA was directly used for the preparation of organozinc compound after the zinc anode was removed from the electrolysis cell. Conversion of ethyl 4-bromobutyrate **1a** into organozinc compound **2a** was determined by GC analysis after quenching of the **2a** with diluted HCl. Conversion of **1a** into **2a** estimated from the analysis was considered as a reactivity index of the electrogenerated zinc (Scheme 2).



Scheme 2

2.3 Cross Coupling Reaction

Organozinc halides were cross-coupled with aryl halides in the same compartment cell. DMF solution (5 ml) of aryl bromide (4 mmol) and Pd(*o*-Tol)₃Cl₂ (0.11 mmol) was added, and the reaction mixture was refluxed for 3 h. The resulting mixture was quenched with HCl solution and filtered. The filtrate was extracted with diethyl ether (50 ml × 3) and the combined organic layers were washed with water (100 ml × 1) and dried over MgSO₄. After evaporation of diethyl ether, the crude product was purified by column chromatography on silica gel with ethyl acetate-hexane (1:4) to give ethyl 4-phenylbutanoate.

3. Results and Discussions

3.1 Electrochemical Preparation of highly Reactive Zinc

Various type of mediator was studied in order to observe the effect of type of mediator on the preparation of highly reactive electrogenerated zinc. Table 1 shows that 2 equivalent of naphthalene is required to convert ethyl 4-bromobutyrate (**1a**) to its corresponding organozinc compound (**2a**) (Entry 2). However, entry 3, 4 and 5 shows that **2a** could be prepared in high yields (87-98%) when phenanthrene, anthracene or pyrene was used as a mediator instead of naphthalene. 0.1 equivalent of those polyaromatics is more enough to provide sufficient reactivity of zinc in order to convert **1a** to **2a**.

Table 1. Effects of various polyaromatic compounds on the reactivity of EGZn/PA^a

Entry	Polyaromatic compound	Equiv.	Conversion (%) ^b
1	None	-	75
2 ^c	Naphthalene	2	98
3	Phenanthrene	0.1	98
4	Anthracene	0.1	91
5	Pyrene	0.1	87

^a Electrolysis was carried out at 60 mA/cm² in 0.1M TEAP-DMF at 0°C. The reactive zinc (6 mmol) was reacted with ethyl 4-bromobutyrate (3 mmol) for 15 min at 70°C.

^b Conversions of **1a** into **2a** were estimated by GC of unreacted **1a**

^c Conversion was calculated at 20 min after **1a** was reacted.

3.2 Cross-coupling of Organozinc Bromide with Aryl Bromide

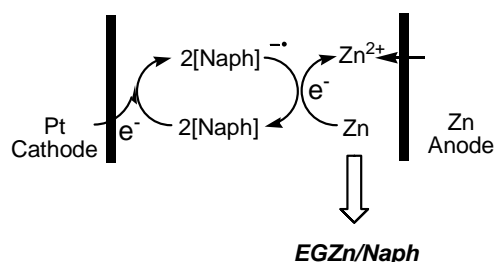
Cross coupling of **2a** with bromobenzene (**3a**) at 70°C for 2 h in the presence of 5 mol% Pd(P(*o*-Tol)₃)₂Cl₂ catalyst gave the corresponding cross-coupled products (**4a**) in 96% isolated yields (Table 1, entry 1). Similar reaction of **2a** with 4-acetylbromobenzene (**3b**), 4-methoxybromobenzene (**3c**) and 1-bromo-3,4-methylenedioxybenzene (**3d**) also took place efficiently to give the corresponding cross-coupled products **4b**, **4c**, and **4d** in 80%, 92%, and 83% isolated yields, respectively (Table 2).

Table 2. Cross-coupling reactions of functionalized alkyl bromides with various aryl bromides using EGZn/Naph

ArI	Product	Yield (%)
C ₆ H ₅ Br (3a)	C ₆ H ₅ (CH ₂) ₃ CO ₂ Et (4a)	96
<i>p</i> -CH ₃ COC ₆ H ₄ Br (3b)	<i>p</i> -CH ₃ COC ₆ H ₄ (CH ₂) ₃ CO ₂ Et (4b)	80
<i>p</i> -CH ₃ OC ₆ H ₄ Br (3c)	<i>p</i> -CH ₃ OC ₆ H ₄ (CH ₂) ₃ CO ₂ Et (4c)	92
(3,4-OCH ₂ O)C ₆ H ₃ Br (3d)	(3,4-OCH ₂ O)C ₆ H ₃ (CH ₂) ₃ CO ₂ Et (4d)	83

3.3 Proposed Reaction mechanism

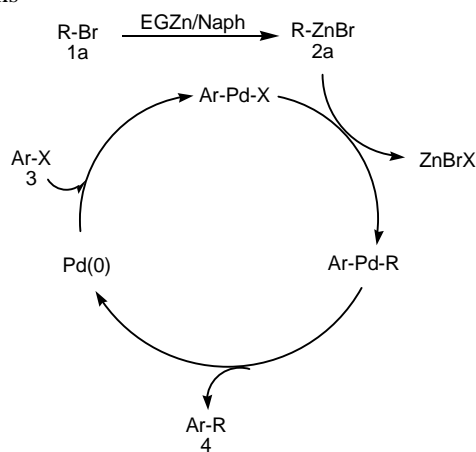
At the cathode, a one-electron reduction of naphthalene molecule readily occurred to give naphthalene radical anion preferentially. The formation of the naphthalene radical anions was shown by the dark green colour that appeared on the surface of the cathode. On the other hand, at the anode, dissolution of the zinc metal occurred to give zinc ions, which were reduced by the naphthalene radical anions to give zero-valence highly reactive zinc, EGZn/Naph (Scheme 3).



Scheme 3

Probable reaction pathways of the present cross-couplings are shown in Scheme 4. Oxidative addition of Pd(0) to aryl halides **3** would give Ar-Pd-X, which undergoes metal exchange reaction with organozinc bromide **2a** to give an intermediate Ar-Pd-R. Reductive elimination of Ar-Pd-R would give the cross-coupling product **4**.

Scheme 4: Probable reaction pathways of cross-coupling reactions



Scheme 4

4. Conclusion

It has been shown that highly reactive electrogenerated zinc (EGZn/PA) could also be synthesized by the use of other polyaromatic compounds such as phenanthrene, anthracene or pyrene as a mediator instead of naphthalene. 0.1 equivalent of those polyaromatics is more enough to convert ethyl 4-bromobutyrate to its corresponding organozinc compound. Additionally, the cross-coupling reaction of the organozinc bromide with various types of aryl bromides gave the product in high yields.

Acknowledgement

This work was supported by a Grant-in-Aids for Scientific Research (B) (No. 11450342 from the Ministry of Education, Science, Sports and Cultures of Japan and The SAGA Grant (Vot: 73704) from Academic Science Malaysia.

References

- Aishah, A. J., Kurono, N. and Tokuda, M. (2001). Facile Synthesis of Ethyl 2-arylpropenoates by Cross-Coupling Reaction Using Electrogenerated Highly Reactive Zinc. *Tetrahedron*. 58(37):7477-7484.
- Hanson, M. V. and Rieke, R. D. (1995). 1,4-Addition of Secondary and Tertiary Alkylzinc Bromides to $\alpha\beta$ -Unsaturated Ketones Without a Copper Catalyst. *J. Am. Chem. Soc.* 117(43): 10775-10776.
- Knochel, P., Rozema, M. J., Tucker, C. E., Retherford, C., Furlong, M. and AchyuthaRao, S. (1992). *Pre and Applied Chemistry*. 64(4), 361-369
- Konochel, P.; Jones, P. (1999). *Organometallic Reagents. A practice Approach*. New York: Oxford University.
- Kurono, N. Sugita, K. Takasugi, S. Tokuda, M. (1999). One Step Cross-coupling Reaction of Functionalized Alkyl Iodide with Aryl Halides by the Use of an Electrochemical Method. *Tetrahedron*. 55(19): 6097-6018
- Tokuda, M., Kurono, N., and Mimura, N. (1996) *Chem Lett*. 1091
- Tokuda, M., Mimura, N., Karasawa, T., Fujita, H. and Suginome, H. (1993). New Preparation of Reactive Zinc Metal by Electrolysis and its Use for a Facile Isoprenylation of Aldehydes and Ketones. *Tetrahedron Letter*. 34(47):7607
- Tokuda, M.; Kabuki, T.; Katoh, Y.; Suginome, H. (1995). Regioselective Synthesis of β,γ -unsaturated Acids by the Electrochemical Carboxylation of Allylic Bromides Using a Reactive-metal Anode. *Tetrahedron Letter*. 36: 3345.
- Zhu, L. and Rieke, R. D. (1991). The Direct Formation of Functionalized Alkyl(aryl)zinc halides by Oxidative Addition of Highly Reactive Zinc with Organic Halides and their Reactions with Acid Chlorides, α,β -unsaturated ketones, and allylic, aryl, and vinyl halides. *J. Org. Chem.* 56(4): 1445-1453.

Oil Water Contact Analysis and Hydrocarbon Saturation Estimation Based on Well Logging Data

Jarot Setyowiyoto¹⁾ and Ariffin Samsuri²⁾

¹⁾ Universitas Gadjah Mada Indonesia, ²⁾Universiti Teknologi Malaysia
Email: j_setyowiyoto@yahoo.com Phone: +60167 285 312

Abstract

Oil water contact analysis, porosity calculation and hydrocarbon saturation estimation is very important stage in reservoir characterization and hydrocarbon reserve estimation. Well logging data studied comprises of gamma ray log, formation density and compensated neutron logs, electric logs i.e. spontaneous potential log, micro spherical focused log, laterolog deep and laterolog shallow,. The research has been done by qualitative and quantitative analysis. Qualitative analysis includes determination of porous zone, sand and shale base line, water bearing formation and hydrocarbon depletion zone. Quantitative analysis includes calculation of formation temperature, mud filtrate resistivity, shale volume, porosity, and water and hydrocarbon saturation. Porosity value is obtained from density log and than corrected by shale volume and hydrocarbon fluid contain. Hydrocarbon saturation is estimated from water saturations which are calculated from true resistivity (R_t), shale volume (V_{sh}) and corrected porosity (ϕ_c) parameters.

As a result, the lithology of the well # JS-35 is interbedded between sandstone and claystone. The hydrocarbon is trapped in ten porous zones that have the reservoir thick vary between 11- 90 feet. Oil water contact occurred in the 2229 feet depth characterized by the rapid change of resistivity, and underneath became oil depletion zone. The highest hydrocarbon saturation encountered in the zone no.1 that is 85.7% with 19 feet thick and formation temperature is 157.4 °F. In the zone have 31.7% porosity, 3% shale volume and 14.3% water saturation. Porosity is affected by clay volume and their density, whereas the hydrocarbon saturation is strongly influenced by porosity and their value of resistivity log.

Key words: oil water contact, porosity, hydrocarbon saturation.

Introduction

Physical properties of rock such as porosity, permeability and saturation are very important in the interpretation of geophysical data especially in an exploration geophysics, well logging, petroleum engineering, and sub surface petroleum geology. Exploration and development an oil field are always aim to finding new hydrocarbon reserve. Oil production from a well is also affected by several factors include hydrocarbon potential and petrophysic properties such as porosity and permeability. This study emphasized only on porosity and hydrocarbon saturation calculation which are estimated from well logging data.

Porosity is define as the ratio of the volume of void or pore space (V_p) to the total of bulk volume (V) of the rock. There are three log-derived porosity measuring tools in common use that are density, neutron and sonic log. Sonic log was less sensitive to borehole and mud cake variations than density and neutron log. So that the density, neutron or density-neutron combination has become the primary source of porosity calculation, displacing the Sonic (Dewan, 1983). Water saturation (S_w) is the important parameter for hydrocarbon saturation (S_h) calculation, $S_h=1-S_w$. Fluid content of a rock formation include water, gas or oil can be predicted by integrated log, especially resistivity log. When hydrocarbon are present in the rock formation, the mud filtrate will replace the oil and gas immediately around the bore hole. The resistivity profile

will show a flush zone with a low to moderate resistivity since filled with mud filtrate, and in the virgin formation with an extremely high resistivity because of the high saturation in hydrocarbon.

Oil water contact analysis is needed for net pay estimation and boundary delineation between water and oil zone. It can be determined from the resistivity profile. The resistivity in the water zone will lower than in the oil zone.

Data and Method

Logging data drom well JS # 35 used in this study comprises of Spontaneous Potential Log (SP), Micro Spherical Focused Log (MSFL), Laterolog Shallow (LLS), Laterolog Deep (LLD), Gamma Ray Log (GR) and Formation Density Log (FDL). The study has been done by qualitative and quantitative analysis. Qualitative analysis includes determination of porous zone, sand and shale base line, water bearing formation, hydrocarbon depletion zone and oil water contact (OWC). Quantitative analysis includes calculation of formation temperature, mud filtrate resistivity, shale volume, porosity, and water and hydrocarbon saturation. Porosity value is obtained from density log and than corrected by shale volume and hydrocarbon fluid contain. Hydrocarbon saturation is estimated from water saturations which are calculated from true resistivity (R_t), shale volume (V_{sh}) and corrected porosity (ϕ_c) parameters. Quantitative analysis is principally based on shally sand interpretation by Darcy and Schlumberger method, include:

1. Determination of formation temperature (T₂):

$$T_2 = \frac{D(BHT - T_1)}{TD} + T_1 \dots\dots\dots(1)$$

Where:

- T₂ = Formation temperature (°F)
- D = Log depth (feet)
- BHT = Borehole temperature (°F)
- TD = Total depth (feet)
- T₁ = Surface temperature (°F)

2. Determination of resistivity of mud filtrate (R_{mf}):

$$R_{mf2} = R_{mf1} \frac{(T_1 + 6.77)}{(T_2 + 6.77)} \dots\dots\dots(2)$$

Where:

- R_{mf2} = Resistivity of mud filtrate at formation (ohm-m)
- R_{mf1} = Resistivity of mud filtrate at surface (ohm-m)
- T₁ = Surface temperature (°F)
- T₂ = Formation temperature (°F)

3. Determination of shale volume (V_{sh}):

$$V_{sh} = \frac{GR_{log} - GR_{min}}{GR_{max} - GR_{min}} \dots\dots\dots(3)$$

Where:

- V_{sh} = Shale volume (%)
- GR_{log} = Gamma Ray Log (API Unit)
- GR_{max} = Gamma Ray maximum (API Unit)
- GR_{min} = Gamma Ray minimum (API Unit)

4. Determination of Porosity:

$$\Phi_D = \frac{\rho_{ma} - \rho_b}{\rho_{ma} - \rho_f} \dots\dots\dots(4)$$

Where:

- Φ_D = Porosity from density log (%)

- ρ_{ma} = matrix density
- ρ_b = bulk density
- ρ_f = fluid density

5. Determination of porosity corrected by V_{sh} content:

$$\Phi_{Dc1} = \Phi - (V_{sh} \times \Phi_{Dsh}) \dots\dots\dots(5)$$

Where:

- Φ_{Dc1} = Corrected porosity by V_{sh} (%)
- Φ = Porosity (%)
- V_{sh} = Shale volume (%)
- Φ_{Dsh} = Log density-derived porosity in shale zone (%)

6. Determination of porosity corrected by hydrocarbon fluid content

$$\Phi_{Dc2} = \Phi_{Dc1} (1 - 0.1 \text{ Shr})$$

$$\text{Shr} = 1 - S_{xo}$$

$$S_{xo} = \frac{I}{R_{xo} \sqrt{\frac{V_{sh}^{(1 - V_{sh}/2)}}{R_{sh}} + \frac{\phi_{Dc1}}{a x R_{mf}}}} \dots\dots\dots(6)$$

Where:

- Φ_{Dc2} = Hydrocarbon fluid corrected porosity (%)
- Φ_{Dc1} = V_{sh} corrected porosity (%)
- Shr = Residual hydrocarbon saturation
- S_{xo} = Water saturation in flush zone
- R_{xo} = Water resistivity in flush zone
- V_{sh} = Shale volume
- R_{sh} = Resistivity in shale zone
- a = Cementation constant
- R_{mf} = Resistivity of mud filtrate

7. Determination of water resistivity (R_w)

Water resistivity can be obtained from water bearing formation.

$$R_w = \frac{R_{mf}xR_t}{R_{xo}} \text{ ohm-m} \dots\dots\dots (7)$$

Where:

- Rw = Water resistivity in uninvaded zone
- Rmf = Resistivity of mud filtrate
- Rt = True resistivity
- Rxo = Water resistivity in flush zone

8. Determination of Water saturation (Sw)

$$S_w = \frac{1}{R_t \sqrt{\frac{V_{sh}^{(1 - V_{sh}/2)} + \phi Dc2}{R_{sh} + a x R_w}}} \dots (8)$$

Where:

- Sw = Water saturation
- Rt = True resistivity
- Vsh = Shale volume
- Rsh = Resistivity in shale zone
- ΦDc2 = Porosity (%), hydrocarbon fluid correction.
- a = Cementation constant
- Rw = Water resistivity

9. Determination of hydrocarbon saturation (Sh)

$$S_h = 1 - S_w \dots\dots\dots (9)$$

Where:

- Sh = Hydrocarbon saturation (%)
- Sw = Water saturation (%)

Result and Discussion

As a result, the lithology of well # JS-35 is interbedded between sandstone and claystone. The hydrocarbon is trapped in ten porous zones that have the reservoir thick vary between 11-90 feet. Oil water contact occurred in the 2229 feet depth characterized by the rapid change in resistivity, and underneath became oil depletion zone. Qualitative analysis of the logs result some data base for the next qualitative analysis, that are:

- GRmin = 50 API
- GRmax = 175 API
- Rsh = 4 ohm-m
- ΦDsh = 0.194
- pf = 1.1

- ρma = 2.65
- BHT = 172 oF
- Rmf = 0.322 @ 80 oF
- T1 = 80 oF
- TD = 2608 feet
- a = 0.81

Hydrocarbon saturation and corrected porosity has been calculated in the tenth zone of sandstone reservoirs (Table1 and Table 2.)

Table 1- Formation evaluation result, log-derived porosity calculation corrected by Vsh.

No.	Depth (feet)	GR Log (API)	T2 (oF)	Rmf (ohm-m)	Vsh (%)	ρB (gram/cc)	ΦD (%)	ΦDc (Vsh) (%)
1	2194	55	157.4	0.17	3	2.13	33.5	32.9
2	2294	62	160.9	0.167	7	2.1	33.5	34.1
3	2320	75	161.8	0.166	15	2.15	32.3	29.4
4	2347	90	162.8	0.165	26	2.18	30.3	25.3
5	2374	80	163.7	0.164	18	2.18	33.5	30
6	2424	90	165.5	0.162	26	2.2	29	24
7	2442	85	166.1	0.162	23	2.29	29	24.5
8	2467	65	167	0.161	11	2.25	25.8	23.7
9	2515	75	168.7	0.159	18	2.25	25.8	22.3
10	2560	55	170.3	0.158	3	2.2	29	28.4

The highest hydrocarbon saturation of 85.7% encountered in the zone no.1 with 19 feet thick and formation

temperature is 157.4 °F. In this zone have 31.7% porosity, 3% shale volume and 14.3% water saturation.

Table 2 - Formation evaluation result, hydrocarbon saturation and porosity calculation corrected by hydrocarbon fluid content

No.	Depth (feet)	$\Phi Dc1$ (%)	Rxo (ohm-m)	Rt (ohm-m)	Sxo	Shr	$\Phi Dc2$ (%)	Sw (%)	Sh (%)
1	2194	32.9	3	10	0.64	0.36	31.7	14.3	85.7
2	2294	34.1	4	9	0.518	0.482	32.5	25.1	74.9
3	2320	29.4	5	30	0.503	0.497	27.9	15.3	84.7
4	2347	25.3	4	10	0.59	0.41	24.3	28.1	71.9
5	2374	30	6	15	0.44	0.56	28.3	20.9	79.1
6	2424	24	4	10	0.611	0.389	23.1	29.4	70.6
7	2442	24.5	4	9	0.616	0.384	23.6	31	69
8	2467	23.7	5	15	0.407	0.593	22.3	27.2	72.8
9	2515	22.3	2.5	10	0.436	0.564	21	33.6	66.4
10	2560	28.4	2	15	0.873	0.127	28	22.8	77.2

The porosity values of the tenth porous zone of sandstone are vary, range 21-32.5%, it is affected by density and clay volume. The porosity generally decreases with increasing in clay volume (Fig.1). Hydrocarbon saturation of the tenth sandstone porous zone are also vary, range 66.4 – 86.7%, influenced by porosity and the value of resistivity. The hydrocarbon saturation increases with increasing in porosity (Fig.2). The resistivity of the hydrocarbon zones are vary, range 9-15 ohm-m

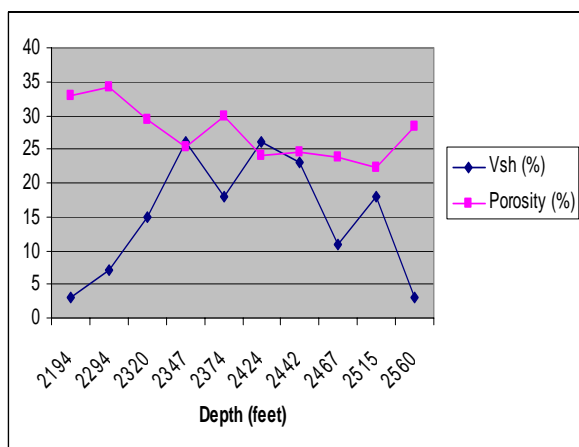


Figure 1- Cross plot of depth vs. porosity and clay content (Vsh).

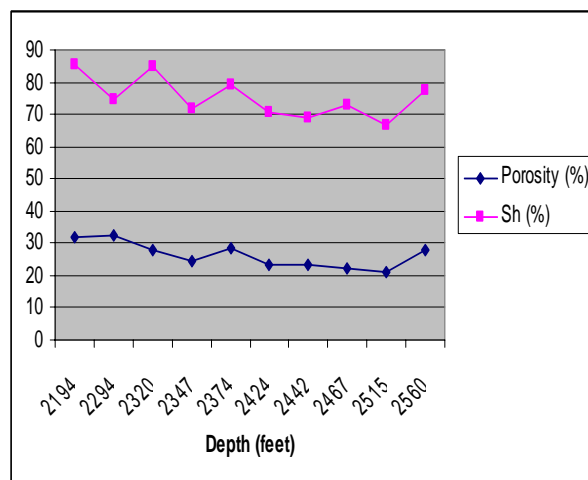


Figure 2- Cross plot of depth vs. porosity and hydrocarbon saturation.

Conclusion

Determination of oil water contact is very important, encountered in the 2229 feet depth and characterized by the rapid change in resistivity value. The porosity values of sandstone porous zone are vary, range 21-32.5%. The porosity generally decreases with increasing clay volume. Hydrocarbon saturation of porous zones are also vary, range 66.4 – 86.7%, primarily affected by their porosities.

Acknowledgement

We thank IRPA (Malaysia Science, Technology and Environment Ministry) project Vot 74058, and PT.CPI for their support data sample.

References

- Amyx**, J.W., Jr., D.M., and Writing R.L., 1996, *Petroleum Reservoir Engineering: Petrophysic*, McGraw-Hill Book Co., New York.
- Chapman**, R.E., 1976, *Petroleum Geology*, Elsevier Scientific Publishing Company, New York.
- Dewan**, J.T., 1983, *Modern Open Hole Log Interpretation*, Scientific Publishing Company, Tulsa, Oklahoma.
- Heysse**, D., 1983, *Open Hole Log Analysis and Formation Evaluation*, Halliburton Logging Service, Houston, Texas.
- Kessler**, C., and Sanders, L., 1989, *Cased-Hole Hydrocarbon Detection in the Permian Basin using the Full-Wave Tool*, SPE-19610 Journal, Texas.
- King**, M., S.Khanshir, M.S., Worthintong, M.H., 1991, *Petrophysic Studies of Core from a Cross-Hole Seismic test site*. The Society of Professional Well Log Analysis, London.
- Klimentos**, T., 1995, *Petrophysical Studies and Seismic Wave Technology: Application in Exploration, Formation Evaluation, and Reservoir characterization*. Research Institute, SPE-29887 Journal, Texas.
- Madon**, M., 2002, *Geological Control on Reservoir Heterogenety*, Technical Paper in PRSS Technology Forum, Petronas Research Proceeding, Malaysia.
- Marion**, D.P., and Pellerin, F.M., 1992, *Acoustic Measurement on Core as a tools for Calibration and Quantitative Interpretation of Log*, SPE-25018 Journal, Texas.
- Schlumberger**, 1989, *Log Interpretation Principles/Application*, Sdhlumberger Wqell Service, Houston, Texas.
- Schon**, J.H., 1996, *Physical Properties of Rocks: Fundamentals and Principles of Petrophysic*, Vol.18, Elsevier, New York.
- Tucker**, M.E., 1991, *Sedimentary Petrology: An Introduction to the Origin of Sedimentary Rocks*, Blackwell Scientific Pub., UK
- Williams**, R.E., 1986, *Formation Evaluation Conference*, 2 nd., PT.Ichtiar Baru-Van Hoeve, Jakarta.

Factors Controlling Sediment Budget at River Mouths in the South Coast of Central Java, Indonesia

Hendra Amijaya^{1*}, Sugeng S. Surjono²

^{1,2} Department of Geological Engineering, Gadjah Mada University
Jalan Grafika 2, Yogyakarta, Indonesia
Tel: +62-274-6491380/-513668, Fax: +62-274-513668
* Email: hamijaya@ugm.ac.id

Abstract

A method to assess the dynamic of sedimentation on river mouth is by analyzing the controlling factors of the sediment budget. Progo, Opak, Bogowonto and Serang Rivers are the main rivers in South Coast of Central Java, which discharge their sand dominated sediment into Indian Ocean. Geomorphologically, those river mouths can be divided into some geomorphological units, which are beach (including beach dune), alluvial plain, lagoon, river and ocean. The beaches surrounding those river mouths clearly show the sedimentation and erosion processes. The dynamic of those processes is indicated by the geomorphological change, such as river mouth shifting, beach line change and dune movement. The budget of sedimentation is depending on the sediment entering the beach from the river, the sediment credit-debit difference between beach and alluvial plain by wind action, sediment credit-debit difference between beach and ocean by wave action, sediment credit-debit difference between beach, river and lagoon, and also the sediment credit-debit difference caused by longshore current. Therefore, the dynamic of the beach is caused by some physical processes with water and wind as main agent of sediment transportation. However, man induced activities such as sand mining also contributes the change of the sediment budget.

Keywords

Sediment Budget, River Mouth, Geomorphology, South Coast of Central Java, Indonesia

1. Introduction

Progo, Opak, Bogowonto and Serang rivers are located in South Coast of Central Java and discharge the flows into Indian Ocean (Fig. 1). The mouths of those rivers show a complex and dynamic sedimentation system. Geomorphological processes with water and wind as agent cause sedimentation and erosion around the river mouths. The dynamics of sediment deposition and erosion on river mouth surroundings are shown by the morphological change, such as river mouth change, lagoon change, beach line change or dune movement.

The dynamics of sedimentation can be studied from sediment budget analysis. Sediment budget analysis counts the amount of sediment which is deposited on the studied area and the amount of sediment leaving the area. The sedimentation budget will be influenced by the deposition and erosion of sediment in every geomorphological units on this beach. In addition to that, man made activities like sand mining contributes the sediment budget of those areas. This paper shows a model to calculate the total sedimentation budget and its application, most of all in the sense of environmental aspect.

2. Approach and Methods

2.1. Sedimentation Budget

Sedimentation budget (or sedimentation balance) analysis basically evaluates the amount of the sediment that entering and leaving a system, in this case river mouth and its surroundings. Based on this analysis, It can be predicted whether there will be an erosion or accretion on the beach near the river mouths. If the budget is zero, then the unit is in stable condition. If the budget is positive, then an accretion or forward movement of the beach occurs. An erosion or backward movement of the beach occurs if the budget is negative.

Bowen and Inman (1966, in Komar, [1]) stated that sedimentation balance is the application of the Mass Conversion Law to the littoral sediment. The grade of the sedimentation balance in a system is influenced by the amount of sediment that entering the system (credit) compared with the sediment that leaving the system (debit).

The sediment credit can come from longshore transport, onshore transport, river supply, beach cliff erosion, wind transport, biogenic deposition, hydrogenic deposition or beach reclamation. The sediment debit can be caused by longshore transport, offshore transport, wind



Figure 1. Location of Bogowonto, Serang, Progo and Opak rivers in South Central Java.

transport, sediment transport into the trench, solution and abrasion of beach material or beach mining. The balance is reflected in the accretion or erosion on the beach.

For the beach that have a dune system on its surroundings, Sherman and Bauer [2] wrote a conceptual model that connected the beach sedimentation budget, dune sedimentation budget and the dune morphology. The interaction of both budget cause a different dune morphology, whether there will be a dune development, dune migration, blowouts and deviation hollows or dune erosion and wash over. A comprehensive study of sedimentation balance calculation can be also seen in the study conducted by Kirk [3] in New Zealand. He made a model for beach-river-lagoon system.

2.2. Methods

Assessment of factors controlling sediment budget at the studied river mouths are basically based on the geomorphological and geological mapping to determine the morphological unit which contributing credit and debit of sediment. Beside field mapping, data of morphological units and their change were compiled from multitemporal topographic maps and remote sensing image (aerial photographs and satellite images).

3. Results

3.1. Observation on the Exogenic Processes

Four most important factor which contribute the formation of morphological unit in the studied area are wave, wind, longshore current and river flow. The intensity of each exogenic factor depends on the season.

In the wet season, the wind is influenced by northwest monsoon wind. This will invite wave from southwest direction. The wave intensity is generally weak and will initiate coastal currents and longshore currents to move to the east direction. Therefore the sediment will also move to the east. At the same time, the river flow rises high and causes many sediment deposition around the river mouth. Because the river flow is stronger than longshore currents, river mouth is always open in this season.

On the contrary, in the dry season the wind is influenced

by southeast monsoon wind. This will generate waves from the south and southeast direction. The wave intensity is generally strong and causes shifting of longshore and coastal currents to the west direction. Therefore sediment move to the west. At this time, river flow has decreased and can not prevent the action by longshore currents on the river mouth. The result of this action is that the river mouth will be closed (or partially closed). Flood will happen if river mouth is still closed and river flow increases due to rain in the early rainy season.

3.2. Geomorphology & Geology

The geomorphology of the coast in South Central Java it self has been described by a lot of authors, for example the classical work from Van Bemmelen [4] and also some paper later such as from Bird and Ongkosongo [5], Soetoto and Widiasmoro [6], Woro and Yuniarto [7], Sutikno [8], Sudarmaji [9] or Setyawan [10].

Basically, five geomorphological units are observed on the river mouths and their surroundings, which are beach (including beach dune), alluvial plain, lagoon, river and ocean. The beach is a plain sand beach, which is bordered with the dune area by the berm. The beach dune is located between the beach and aluvial plain. The form of the dunes are mostly transversal and parabolic. The dunes are mostly oriented to the northwest direction. This orientation shows us that the wind come dominantly from the southeast direction. The lagoon located at the west side of the river mouth and oriented to the northwest direction, same with the orientation of the beach. The lagoon was formed because the river flow eroded the west side of the river mouth. The longshore current, therefore the forming of sand spit, is the cause of the formation of this lagoon. From the sand spit morphology at the river mouth it also can be concluded that the nett longshore sand transportation is into west direction.

It should be noted, that the morphological change of around of this river mouth happens oft. Multitemporal aerial photography and topographic maps observation showed that the river mouth moved at a certain time to the west direction, and then shifted back to the original position. This change is especially influenced by longshore current and the river flow debit, which are also influenced

by the seasons as mentioned above.

However, every river mouth has special characteristics. Around the mouth of Serang and Progo river, the geomorphological units that seems very important to be considered for detail analysis of sediment budget are beach, beach dune, lagoon and alluvial plain. At Opak river mouth, lagoon was not formed even when the the river flow was shifted to the west direction. This could be influenced by the high flow intensity of Opak river and also less sediment was brought by longshore current because limited sediment source available. On Bogowonto river mouth, there are two river flows, which are Bogowonto river itself and Pasir river, a small river located near the river mouth. Lagoon was not formed at this area. The flow intensity of the rivers are very low so that the flows were not able to build a westward shifting even the river mouth were closed by sand spit.

Rahardjo et al [11] stated that the lithology of the coast in southern Yogyakarta composed of the young deposit of Merapi Vulcano in form of aluvial deposit that consist of pebble, sand, silt and clay. Generally, the sediment that is deposited on the beach around those river mouths have a sand size. Some pebble and clay size sediment can be founded on the area direct on the side of the river. Sand size sediment dominate also the beach dune.

4. Discussion

4.1. Factors Controlling Sediment Budget

Based on the characteristics of those river mouths, the sedimentation budget is generally controlled by some factors which are :

- the sediment entering the beach from the river (S)
- the sediment credit-debit difference between beach and dune by wind transportation (K0-K1)
- the sediment credit-debit difference between dune and alluvial plain by wind transportation (D0-D1)
- the sediment credit-debit difference between beach and ocean by wave action (P0-P1)
- the sediment credit-debit difference between beach, river and lagoon (R0-R1 and G0-G1) or river and beach (G0-G1)
- the sediment credit-debit difference caused by longshore current (L0-L1).

To calculate the sedimentation budget (N) in the system surroundings the Serang and Progo River mouths, five morphological units should be considered. Those units are beach, beach dune, aluvial plain, lagoon, river and ocean (Fig. 2). Therefore the general formula for the sedimentation budget model is:

$$N = S+(G0-G1)+(R0-R1)+(K0-K1)+(D0-D1)+(P0-P1)+(L0-L1).$$

For Opak river mouth, the lagoon does not exist (Fig.3). Therefore the sediment budget can be calculated as follows:

$$N = S+(G0-G1)+(K0-K1)+(D0-D1)+(P0-P1)+(L0-L1).$$

At Bogowonto river mouth, there are two sources of sediment input from river flows (Fig.4). The budget can be formulated as:

$$N = (S1+S2)+(G0-G1)+(K0-K1)+(D0-D1)+(P0-P1)+(L0-L1).$$

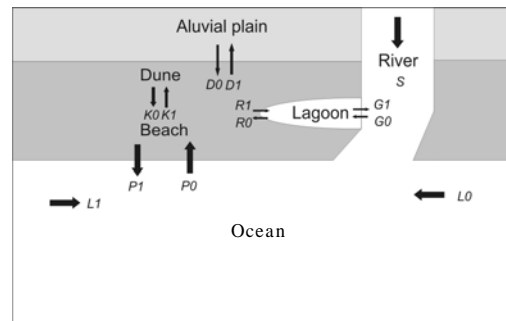


Figure 2. Factors controlling sediment budget on the Serang & Progo river mouth.

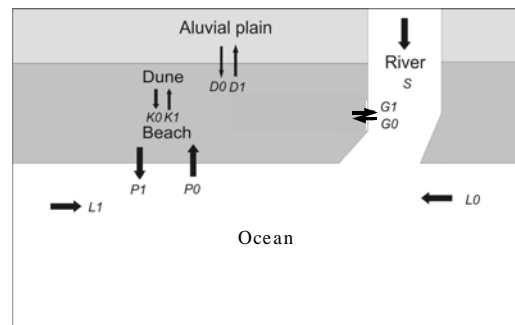


Figure 3. Factors controlling sediment budget on Opak river mouth.

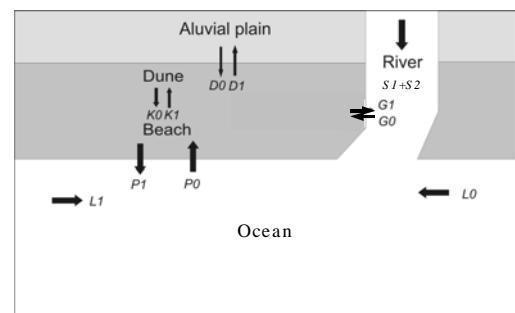


Figure 4. Factors controlling sediment budget on Bogowonto river mouth

4.1. Application of Sediment Budget Analysis

The advantage from this analysis is that the future geomorphological and geological condition of the coast in a definite time can be predicted. The way of handling of the coast can be stated from this analysis. Some examples

of the application are:

- **Land use planning**

Coast can be developed for harbour, fishery, agriculture, tourism, sport, etc. On the coast of South Central Java, some development have been planned by local government. Important factor that some times is not exactly observed is the future condition of the area caused by nature process, such as morphological change caused by sedimentation process on the beach. In the report of South Java Flood Control Survey Project [12], it is stated that in around of 60 years, the accretion of the beach direct on the west side of Progo River mouth reaches 100 m and on the east side is around 30 m eroded.

Land use planning that not consider such process is useless because what that have been planned or builded can not be used again even destroyed in a short time. Such cases are already happened in some tourism site at the beach in South Yogyakarta. Some of the facilities for tourism can not be used again, because the facilities are eroded, buried or because of beach accretion, the facilities become too far away from the beach line. The prediction of the dynamic of the beach using sedimentation budget analysis can be very useful because the land use can be planned exactly and when it is needed, the sedimentation process can be controlled to protect some area.

- **Flood protection/control and water resources planning**

A typical natural disaster that of happens in a river mouth is flood. For example on the Bogowonto River mouth, flood happens yearly and sometimes flooded some villages around. The river water floods the area around the river mouth because of the river mouth are closed by the sand spit. Some flood control structures and actions have been planned to overcome this problem.

The flood control will be much better, if the flood controlling are done not only on the flood site, but based on a regional observation, especially all the factors that may have an effect to the forming of sand spit. The regulation of the sediment credit and budget can be done based on the sedimentation budget analysis. For example to calculate how much exactly sediment that allowed to enter the river mouth or how much sediment that allowed to be transported by longshore current without causing a river mouth closing. From this analysis than the next flood control activities can be exactly determined.

This activities have also an advantage for the water resources planning, especially for irrigation planning. If the sedimentation at the river mouth can be controlled, that means that the irrigation for agricultural or fishery need can be exactly predicted, without any doubt of missplanning because of the flood.

- **Sand mining planning**

Some type of beach sand can be mined to be used, for example for building material and mainly to be extracted if the sand contains economic mineral. Alwi [13] stated that the beach sand at south coast of Central Java have Fe-minerals contain, especially magnetit and ilmenit, which are

composed 14% of the sand. At some part of this area (most of all between Serang and Bogowonto rivers), the beach sand is already mined to be extracted.

The sand mining could give a bad impact to the beach, if the mining is not gut planned. As what already happened, some abandoned mining area caused environmental problems, although the reclamation is already made. For example some beach abrasion/erosion occur. The beach abrasion or erosion is occuring because the debit of sediment is still bigger than the credit. That makes the sedimentation is not in balance. It is clear that sand mining contributes to a very high value of debit and reclamation as its credit and also another credit factors cannot catch the amount of the debit of sediment. Therefore the balance is negative.

To keep the sedimentation in balance, the sedimentation budget should be calculated. How much sand can be mined each year can be known from this analysis. To calculate the budget, the factor of sediment credit and debit between the mining and the beach can be added to the formula. The model can be seen in Figure 5.

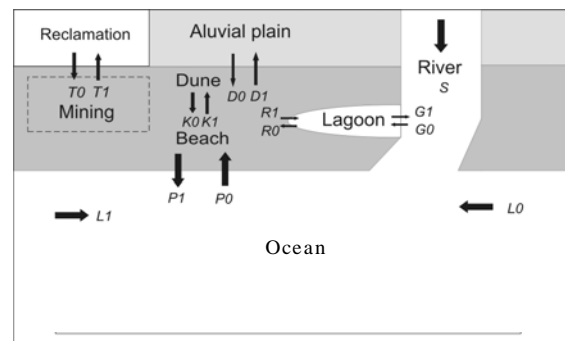


Figure 5. For the beach mining planning need, the factor T_0 and T_1 caused by mining and reclamation can be added into the model for sediment budget calculation (for example Progo river).

- **Nature preservation**

The coast of South Central Java, especially in Yogyakarta, have a unique characteristic. Its famous uniqueness is the beach dune. This area is not only very important from the point view of science, but also very attractive for a tourist destination. This area has been proposed to be a nature conservation. The idea to make this area as a nature conservation area is useless, if in the future the area that will be conserved changes. The anthropogene factors that endangered the preservation can be limited by a certain rule, but the nature process itself should be also controlled so that the preservation can be attained.

Since the dune formation is very effected by the sedimentation process on its surroundings, the favourable degree of sedimentation balance should be known. The balance of sedimentation on the beach and on the dune itself should be comprehend to make the dune always preserved. For example if the both budget are negative, then a dune erosion and washover will be happened. The sedimentation budget analysis is needed to determine a plan

to control the interaction of sedimentation of each unit. Therefore the supply of sand is always sufficient to form the dunes, and also sufficient to preserve dune.

5. Conclusions

Sedimentation budget analysis is useful to study the dynamic of an area around river mouth, because in this area, the interaction of some sedimentation processes are intensively occurred. Principally, the geomorphological conditions and exogenic factors are controlling the sediment budget. Those geomorphological units are beach, beach dune, alluvial plain, lagoon, river and ocean. The formation of those geomorphological units and their change are influenced by exogenic factor such as wave, wind, longshore current and river flow. The analysis is actually based on the budget calculation of sediment credit and debit between each morphological unit.

The analysis is very useful for some need, for example for planning and development of the coastal area, because it provide an overview of the dynamic of the coast. From this knowledge the future condition of the coast can be predicted. The sedimentation budget data is very useful to make a comprehensive coastal planning.

Acknowledgment

Both authors wish to thank all the team member of „Toward Sustainable Coastal Development in The South of Central Java: A Geological Assessment for Coastal Management“ research project conducted by Department of Geological Engineering, Gadjah Mada University, Yogyakarta, Indonesia for the support.

References

- [1] Komar, P.D., 1976. *Beach Processes and Sedimentation*, Prentice-Hall, Inc., New Jersey.
- [2] Sherman, D.J., and Bauer, B.O., 1993. Dynamics of Beach-Dune Systems, *Progress in Physical Geography* 17, 4, pages 413-447.
- [3] Kirk, R.M., 1991, River-Beach Interaction on Mixed Sand and Gravel Coasts: A Geomorphic Model for Water Resources Planning, *Applied Geography*, Vol.11/1991, pages. 267-287.
- [4] Van Bemmelen, R.W., 1949. *The Geology of Indonesia*, Vol. IA, 2nd Ed., Martinus Nijhoff, The Hague.
- [5] Bird, E.C.F., and Ongkosongo, O.S.R., 1980. *Environmental Changes on The Coasts of Indonesia*, The United Nations University, Tokyo.
- [6] Soetoto and Widiasmoro, 1987. Concepts of Coastal Region Development in Southern Yogyakarta Province Based on Geological Data Through Landsat Imagery and Aerial Photographs Interpretations, *Proceeding of Regional Conference on Geology, Mineral and Hydrocarbon Resources of South East Asia*, IAGI, Jakarta, pages. 413-423.
- [7] Woro, S. and Yunianto, T., 1988. Geomorphological Analysis for Assessment of Coastal Recreation Sites in The Coastal Area of Trisik, Kulon Progo Regency, Yogyakarta Province, *The Indonesian Journal of Geography*, Vol. 18, No. 56, December 1988, pages. 43-54.
- [8] Sutikno, 1996. Geomorphology of Yogyakarta Area and Its Surrounding Proposed as A Geomorphological Field Laboratory, *The Indonesian Journal of Geography*, Vol. 28, No. 71, June 1996, pages. 1-10 .
- [9] Sudarmaji, 1995. Assessment of Groundwater Resources at The Beach Ridges of The Kulon Progo District, Yogyakarta, Indonesia, *The Indonesian Journal of Geography*, Vol. 27, No. 69, June 1995, pages. 61-74.
- [10] Setyawan, W.B., 1999. Pantai Pasir di Wilayah Pesisir Yogyakarta, *Prosiding Seminar Nasional Sumberdaya Geologi*, Jurusan Teknik Geologi FT UGM, Yogyakarta, pages 239-243
- [11] Rahardjo, W., Sukandarrumidi dan Rosidi, 1995. *Geological Map of Yogyakarta, Jawa*, Directorate of Geology, Republic of Indonesia, Bandung.
- [12] South Java Flood Control Survey Project, 2000, *Coastal Report*, February 2000, Direktorat Jendral Pengembangan Pedesaan, Departemen Pemukiman dan Pengembangan Wilayah RI. .
- [13] Alwi, M.J., 1999, Manajemen Pengelolaan Pesisir Selatan DIY dan Jawa Tengah serta Prospek Pembangunan Pabrik Besi Baja Terpadu, *Prosiding Lokakarya Eksplorasi, Eksploitasi dan Pengolahan Sumberdaya Mineral*, Jurusan Teknik Geologi FT UGM, Yogyakarta, pages 1 – 11.

Decolourization of Azo Dye Direct Blue 15 Using Batch Culture of *Klebsiella* sp.

**Chun Shiong Chong, Zaharah Ibrahim, Madihah Md Salleh,
Noor Aini Abdul Rashid, Adibah Yahya, Wui Jin Wong**

Department of Biology, Faculty of Science
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
Tel: +60-7-5532528, Fax: +60-7-5566162, E-mail: tccshiong80@yahoo.com

Abstract

Nowadays, the development of textile industries are mushrooming and the pollution resulting from the dye-wastewater is alarming. Colour is recognized as the first contaminant from the dye containing wastewater. The present study reports the decolourization of Direct Blue 15 by *Klebsiella* sp., originated from textile wastewater plant. The important parameters, including co-substrates, aeration, starting pH of culture and temperature for decolourization, were investigated. Mechanism of colour removal either by bioadsorption or biodegradation was determined. Amongst the co-substrates, glucose (0.2% w/v) was the preferred co-substrate for decolourization of Direct Blue 15. The optimal decolourization occurred under partial anaerobic condition at starting pH of about 6.0. Decolourization of Direct Blue 15 was most efficient at 45°C. Under the optimum conditions, it was found that the COD removal was 53.6%. Dye decolourization (81.9%) was successfully achieved within 24 hours; 0.03% was due to the bioadsorption and 81.87% was due to the biodegradation by *Klebsiella* sp., while the product of degradation of azo dye, sulfanilic acid, was detected using HPLC.

Keywords: Direct Blue 15, decolourisation, azo dye, *Klebsiella* sp., HPLC

1. Introduction

Azo dyes are extensively used in many fields such as textile industries, leather tanning industries, food industries, cosmetics industries and plastics industries [1,2]. The annual production of azo dyes, the largest group of dyes was estimated about 1 million tons [1].

Azo dyes are recalcitrant xenobiotics, thus, they are resistant to light, heat, oxidizing agent and biodegradation [3]. Thus, many of the dyes, especially synthetic azo dyes are difficult to remove by conventional waste treatment methods as they are stable due to their complex aromatic molecular structures. During the past two decades, some physico-chemical treatment methods have been suggested, however only a few of them are accepted due to low efficiency, high cost and inapplicability to a wide range of dyes [4]. Recently, Department of Environment (DOE), Malaysia maybe imposed colour as one of parameter under the regulatory lead to the requiring of more efficient and cost-effective treatment methods.

The use of microorganisms to decolour and biodegrade synthetic azo dyes from the industrial effluents offers appreciable advantages where it is relatively inexpensive and may detoxify the compounds via mineralization process [2]. It was reported that under anaerobic condition, azo dyes are

readily reduced into their intermediates which are normally colourless aromatic amine compounds, cannot be further degraded under anaerobic degradation [5]. Aromatic amines on the other hand can be further degraded under aerobic condition [6].

In the present study, Direct Blue 15, a water soluble azo dye (Figure 1) was selected for carrying out microbial decolourization studies. Besides that, optimization of decolourization in different parameters, such as co-substrates, aeration, starting pH of culture and temperature, was also carried out. To confirm if decolourization process was due to the reduction of azo dyes, HPLC analysis was performed.

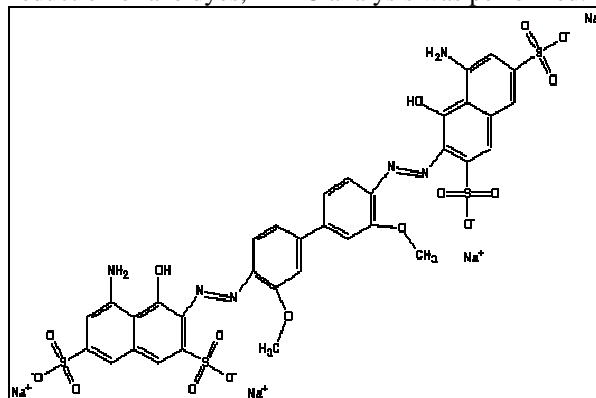


Figure 1. Structure of C.I. Direct Blue 15 [7]

* Corresponding Author. E-mail: zaharah@bio.fs.utm.my
Tel: +60-7-5534122, Fax: +60-7-5566162

2. Materials and methods

2.1 Chemical

Direct Blue 15 (C.I 24400) and sulfanilic acid (99 %) were purchased from Sigma and Fluka, respectively. All other chemical used throughout this study were of analytical grade.

2.2 Microorganism

Klebsiella sp. was isolated from a textile wastewater treatment using serial dilution method. This strain had the ability to remove some pure dyes, such as Orange II, Direct Blue 15, SF Yellow 3RS, SF Red 3BS, SF Yellow EXF and SF Black EXA, under partial anaerobic condition.

2.3 Growth medium

Filtered textile wastewater was used as medium in this study. This medium was supplemented with carbon source (sodium acetate, glycerol or glucose) and azo dye Direct Blue 15 (0.01g/L) from their respective filter sterilized stock solutions.

2.4 Optimization of decolourization

2.4.1 Effect of co-substrates

Three types of co-substrates had been used to determine their effect on decolourization process; these were glucose, glycerol and sodium acetate. Inoculum (10% v/v) was added into the textile wastewater medium containing Direct Blue 15 (0.01g/L). Decolourization under partial anaerobic condition was carried out by filling up the universal bottle, and then incubated under 37°C without shaking at pH 7 for 40 h. From these experiments, the most preferred co-substrate was used in subsequent experiments.

2.4.2 Effect of aeration

Inoculum (10% v/v) was added into the textile wastewater containing Direct Blue 15 (0.01g/L) and selected co-substrate (0.1% w/v). Decolourization under partial anaerobic condition was carried out by filling up the universal bottle and incubated at 37°C. For aerobic condition, the cultures were shaken at 100 rpm and 150 rpm at 37°C and incubated in a period of 40 h.

2.4.3 Effect of temperature

A range of temperature (27°C - 60°C) was used to study the effect of the temperature on azo dye decolourization. Inoculum (10% v/v) was added into textile wastewater medium containing Direct Blue 15 (0.01g/L) and selected co-substrate (0.1% w/v) at pH 7 and incubated under partial anaerobic condition over a period of 40 h.

2.4.4 Effect of pH

Effect of pH for decolourization of textile wastewater medium containing Direct Blue 15 (0.01g/L) was studied over a pH range of 5 to 10 by adjusting pH with HCl and NaOH. Incubation was carried out under optimized temperature, using selected co-substrate under partial anaerobic condition over a period of 40 h.

2.4.5 Effect of co-substrate concentration

Effect of selected co-substrate concentration for decolourization of textile wastewater medium was studied. Different concentrations of glucose, ranging from 0.1% to 0.5% (w/v) were used.

2.5 Batch culture experiment under optimized condition

Klebsiella sp. (10% v/v), at its active phase was transferred into medium for decolourization experiment. Incubation was done under optimized conditions, that is pH 6, glucose concentration (0.2% w/v), temperature (45°C) under partial anaerobic condition. The parameters such as absorbance of dye, bioadsorption, COD and dye intermediates were monitored.

2.6 Analytical methods

2.6.1 Decolourization study

Absorbance of Direct Blue 15 measurements in medium were performed in a CARY 100 Bio spectrophotometer in the UV-Visible range. Colour was measured at the dye's optimum wavelength (584 nm). For this purpose, samples were centrifuged at 10 000 rpm for 15 min and absorbance value of supernatants were determined. The decolourization efficiency was expressed as the following Equation (1).

$$\text{Decolourization (\%)} = \frac{(I - F)}{I} \times 100 \quad (1)$$

where I = initial absorbance and F = absorbance of decolourized medium.

2.6.2 Bioadsorption study

The pellet from the centrifuged sample was suspended in 10 mL of distilled water. This was then vortexed and filtered using cellulose acetate membrane filter (0.2 µm) and the filtrate (liquid phase) was measured at 584 nm for the bioadsorption study. The decolourization efficiency due to bioadsorption was expressed as the following Equation (2).

$$\text{Decolourization efficiency by bioadsorption (\%)} = \frac{L}{I} \times 100 \quad (2)$$

where I = initial absorbance and L = absorbance of filtrate (liquid phase)

2.6.3 High Performance Liquid Chromatography (HPLC) and COD analysis

HPLC (Agilent 1100) equipped with a UV detector was used to determine the presence of the sulfanilic acid (dye biodegradation product) in the samples. The samples were eluted isocratically using a C_{18} reversed phase column. The mobile phase was methanol/ H_3PO_4 /water (50.0: 0.6: 49.7 by vol.) at 0.7 ml/min and the detection wavelength was set as 254 nm. COD values were measured according to the standard method [8].

3. Results and discussion

3.1 Effect of co-substrates

Different types of co-substrates namely sodium acetate, glycerol and glucose were used to determine their effects on decolourization of Direct Blue 15.

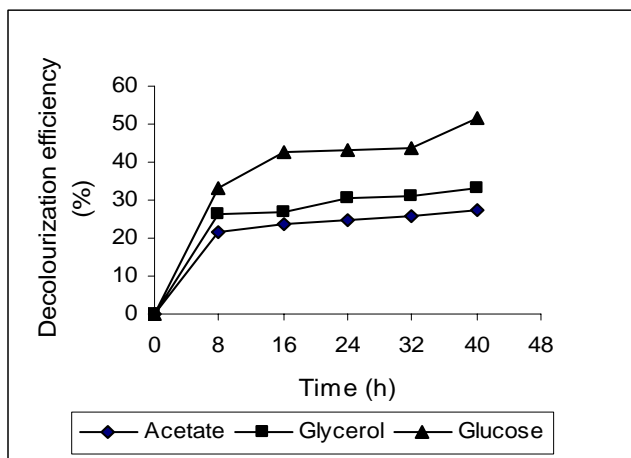


Figure 2. Effect of co-substrates on decolourization of Direct Blue 15 in wastewater medium at 37°C and pH 7

The results presented in Figure 2 indicated that glucose was the most preferred co-substrate since the presence of glucose in the medium showed highest decolourization efficiency (51.4%), compared to glycerol (33.0%) and sodium acetate (27.2%). According to Delée *et al.* (1998), lactose, glucose, glycerol and yeast extract had been reported to be essential co-substrate to improve the decolourization efficiency [5]. A study carried out by Nigam *et al.* (1996) demonstrated that glucose gave the best result in terms of colour removal efficiency (82%) and this was in good

agreement with the results obtained in this study [9]. Subsequently, glucose was used as co-substrate for the following experiment.

3.2 Effect of aeration

Different aeration rates ranging from 0 rpm to 150 rpm was used to measure its effect on decolourization. It was found that non-aerated (0 rpm) condition showed highest decolourization efficiency.

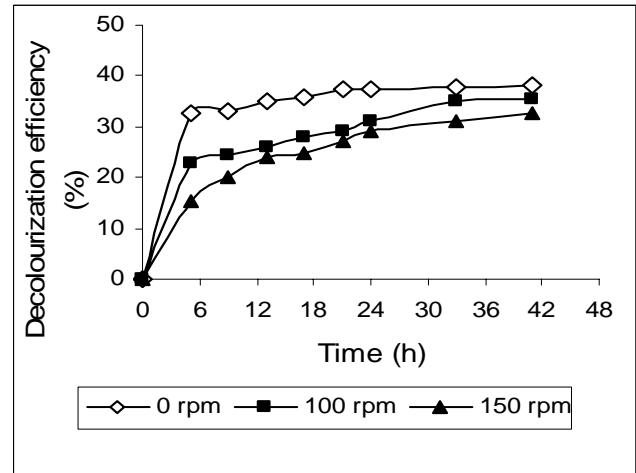


Figure 3. Effect of aeration on decolourization of Direct Blue 15 in wastewater medium containing glucose (0.1% w/v) under 37°C at pH 7

Figure 3 also indicated that the higher aeration rate, the lower decolourization rate. This means the reduction of Direct Blue 15 occurred in the absence of oxygen. This was due to the fact that the presence of oxygen would have a significantly negative effect on the dye reduction, where oxygen may act as high-redox-potential electron acceptor, thus, it would accept the electron from the electron donors rather than azo dye [10]. A wide range of dyes were reported to be decolourized anaerobically [5]. Under anaerobic conditions, azo dyes are used as terminal electron acceptor in electron transport chains and breakdown of azo bond, resulted in decolourization of wastewater [11]. This process however, was inhibited by the presence of oxygen, as oxygen is thermodynamically favorable oxidizing agent [12].

3.3 Effect of temperature

Results obtained for the effect of temperature on decolourization is shown in Figure 4.

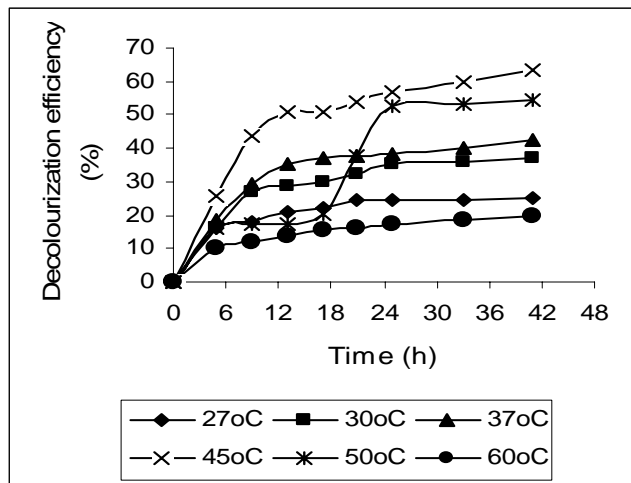


Figure 4. Effect of temperature on decolourization of Direct Blue 15 in wastewater medium containing glucose (0.1% w/v) under partial anaerobic condition at pH 7

The decolourization of Direct Blue 15 was most efficient at 45°C, with 63.3% of colour removal, followed by 50°C, 37°C, 30°C and 27°C. As indicated in Figure 4, at 60°C, the colour removal percentage was only 19.9%. The decline of microbial activity on decolourization could be due to the effects on enzyme denaturation that led to inactivation of the enzyme [13]. Similarly, Pearce *et al.* (2003) also reported that the decline decolourization efficiency might be attributed to loss of cell viability and denaturation of azoreductase enzyme [10].

Temperature influences the metabolic activities of the microbial population. It was noted that temperatures below the optimum typically have a more significant effect on growth rate than temperatures above the optimum [14]. In this study, the optimum growth temperature for *Klebsiella sp* was 37°C while the optimum temperature for colour removal was 45°C. Hence, the starter culture for the bacteria were grown at 37°C while the experiments for colour removal were carried out at temperature 45°C.

3.4 Effect of pH

The optimum pH for decolourization in this study were pH 5 and pH 6, in which decolourization efficiency was 49.9%.

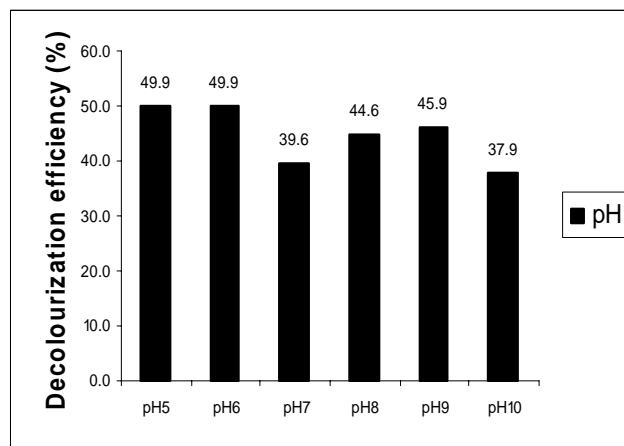


Figure 5. Effect of pH on decolourization of Direct Blue 15 in wastewater medium containing glucose (0.1% w/v) under partial anaerobic condition at 45°C

Klebsiella sp. performed better under slightly acidic condition. It was reported that the decolourization rate increased nearly 2.5-fold as the pH was raised from 5.0-7.0. However the decolourization rate becomes insensitive to pH ranging from 7.0-9.5 [10]. Amongst both optimum pH, pH6 was used for subsequent experiment as it was near to neutral and suitable for industrial application.

3.5 Effect of glucose concentration

Different initial glucose concentrations of glucose (0.1-0.5% w/v) in wastewater medium were used to determine their effect on decolourization efficiency of Direct Blue 15.

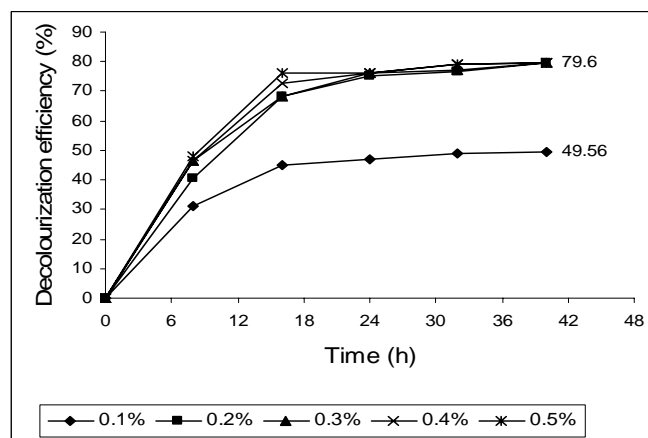


Figure 6. Effect of glucose concentration on decolourization of Direct Blue 15 in wastewater medium under partial anaerobic condition at 45°C and pH 6

The results presented in Figure 6 indicated that decolourization efficiency increased from 0.1% (w/v) to 0.2% (w/v) of glucose. However, further increased in glucose concentration up to 0.5% (w/v) had no significant effect on

decolourization rate of dye. Thus, 0.2% (w/v) of glucose in wastewater medium was optimum to achieve efficient decolourization. According to Sponza and Isik, (2002), a sequential anaerobic/aerobic system, which the medium supplemented with glucose (as 3000 mg/L COD), was efficient to remove colour (Reactive Black 5) up to 98% in anaerobic stage [12].

3.6 Batch culture experiment under optimized condition

Under optimized condition, the results showed the decolourization efficiency was 81.90%.

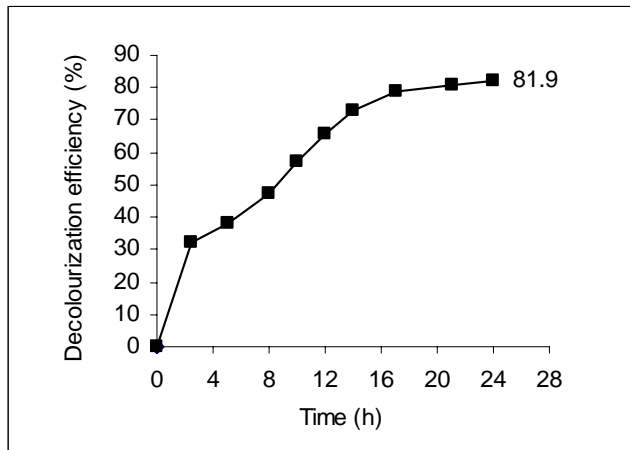


Figure 7. Decolourization of Direct Blue 15 under optimized condition

It should be noted that decolourization process might occur by bioadsorption or/and biodegradation. Thus, it was interesting to determine the percentage colour removal that contributed by bioadsorption and biodegradation. Results indicated that the decolourization due to bioadsorption and biodegradation were 0.03% and 81.87%, respectively. Pearce *et al.* (2003) reported that adsorption of dye onto the biomass was the simplest mechanism to remove colour [10]. However, as similar with the physical adsorption mechanisms for colour removal, this process would become a stage, called saturated, and hence, it was not a good process for long term of colour removal. On the other hand, colour removal via biodegradation was preferred as it remove colour by degrading the dyes into simpler compounds, which is less toxic than the parent compounds.

To further confirm that decolourization had occurred by biodegradation process, HPLC analysis was used to detected products of degradation. It is known that sulfanilic acid was one of the products for degradation of Direct Blue 15. Hence, the presence of sulfanilic acid in the decolourized sample can confirm that decolourization process occurred by biodegradation.

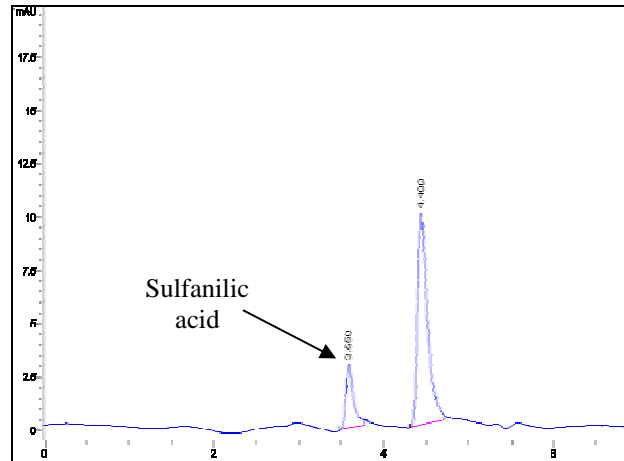


Figure 8. Chromatogram for sample after 24 h of incubation time under optimized condition

As shown in Figure 8, sulfanilic acid peak was present at retention time 3.55 minutes. Therefore, it could be concluded that biodegradation was one of the mechanisms involved in decolourization of Direct Blue 15.

3.7 Removal of COD

A total amount of 53.6% of COD was successfully removed after 24 hours of incubation time. The initial COD level in the wastewater medium was 1815 ppm. During the first 5 hours of incubation time under the optimum condition, the COD level decreased by 51.9%, from 1815 ppm drastically dropped to 873 ppm. However, after that, the COD reduction rate became slow. This may be due to the accumulated of the aromatic amines which are anaerobic recalcitrant inhibited the reduction of the COD level [15]. Furthermore, the change of pH in wastewater medium into more acidic condition might possible affect the reduction of COD.

4. Conclusion

The optimum parameters for Direct Blue 15 decolourization, such as co-substrates, aeration, starting pH of culture and temperature, were of glucose (0.2% w/v), under partial anaerobic condition, pH6 and 45°C respectively. Under these optimized conditions, 81.9% of colour removal was successfully achieved. It was found that both bioadsorption and biodegradation were involved in the colour removal process. Sulfanilic acid as one of the biodegradation product was successfully detected by using high performance liquid chromatography. Under the optimized condition, the COD removal was 53.6%.

Acknowledgements

This work was funded by Malaysian Ministry of Science, Technology and Innovation under IRPA, vote 74053.

References

- [1] Stolz, A. 2001. Basic and Applied Aspects in the Microbial Degradation of Azo Dyes. *Appl Microbiol Biotechnol* 56:69-80.
- [2] Forgacs, E., Cserháti, T., Oros, G. 2004. Removal of Synthetic Dyes from Wastewaters: A Review. *Environ Int* 30:953-971.
- [3] Manu, B, Chaudhari, S. 2001. Anaerobic Decolourisation of Simulated Textile Wastewater Containing Azo Dyes. *Bioresource Technol* 82:225-231.
- [4] Banat, IM., Nigam, P., Singh, D., Marchant, R. 1996. Microbial Decolourisation of Textile Dye Containing Effluents: A Review. *Bioresource Technol* 58:217-227.
- [5] Delee, W., O'Neill, C., Hawkes, F.R., Pinheiro, H.M. 1998. Anaerobic Treatment of Textile Effluents: A Review. *J Chem Technol* 73:323-335.
- [6] Supaka, N., Juntongjin, K., Damronglerd, S., Delia, M-L., Strehaiano, P. 2004. Microbial Decolourization of Reactive Azo Dyes in A Sequential Anaerobic-aerobic System. *Chemical Engineering Journal* 99:169-176.
- [7] Alden, C.J., Boorman, G.A., Bristol, D.W., Dunnick, J.K., Eustis, S.L., Goehl, T.J., Griesemer, R.A., Haseman, J.K., Jokinen, M.P., Rao, G.N., Walters, D.B., Witt, K.L. 1992. Toxicology and Carcinogenesis Studies of C.I. Direct Blue 15 in F344/N Rats." *National Toxicology Program Reports* 14.
- [8] Water Analysis Handbook. (2002). 4th edition. HACH Company.
- [9] Nigam, P., McMullan, G., Banat, I. M., Marchant, R. 1996. Decolourization of Effluent from the Textile Industry by A Microbial Consortium. *Biotechnology Letters*. 18:117-120.
- [10] Pearce, C.I., Lloyd, J.R., Guthrie, J.T. 2003. The Removal of Colour from Textile Wastewater using Whole Bacterial Cells: A Review. *Dyes and Pigments* 58:179-196.
- [11] Isik, M. 2004. Efficiency of Simulated Textile Wastewater Decolorization Process Based on the Methanogenic Activity of Upflow Anaerobic Sludge Blanket Reactor in Salt Inhibition Condition. *Enzyme and Microbial Technology* 35 (5):399-404.
- [12] Sponza, D.T. and Isik, M. 2002. Decolorization and Azo Dye Degradation by Anaerobic/aerobic Sequential Process. *Enzyme and Microbial Technology* 31:102-110.
- [13] Shuler, M.L., Kargi, F. 2nd edn. 2002. *Bioprocess engineering basic concepts*. Prentice Hall, NJ. USA.
- [14] Mul Kerrins D., Dobson D.W., Colleran.E. 2004. Parameters Affecting Biological Phosphate Removal from Wastewaters, *Environment International* 30(2):249-259.
- [15] O'Neill, C., Hawkes F.R., Hawkes, D.L., Esteves, M.S. and Wilcox, S.J. 2000. Anaerobic-aerobic Biotreatment of Simulated Textile Effluent Containing Varied Ratios of Starch and Azo Dye." *Water Research* 34(8):2355-2361.

Microbial Fuel Cell Operating on Cellulose for Electricity Generation

Aslizah, M.A , Adibah, Y. * , Zaharah, I.

Biology Department, Faculty of Science,
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +607-5532528, Fax: +607-5566162, E-mail: aslizah_n@yahoo.com

Abstract

A Microbial Fuel Cell (MFC) was optimized in terms of operational parameters for electricity generation by using cellulase-producing bacteria. Two strains coded as strain Bb and P9 have been used to study their potential to generate electricity during cellulose metabolism. The performance of MFC was analyzed based on their potential (open circuit) and current (close circuit) generated by both strains respectively. The ability of these isolates to use cellulose as their sole source of carbon was indicated by the production of cellulases (CMCase, FPase and β -glucosidase) during growth. This work also demonstrated that cellulases production (cellulose metabolism) was directly related with electricity generation by both strains. Strain Bb and P9 resemble each other by their optimum temperature (37 °C) for growth and electricity production. However, they were distinguished by their respective optimum pH (6.0 and 7.0), carbon (native cellulose and empty fruit bunch) and nitrogen sources (casamino acid + peptone and yeast extract + peptone). At the optimum conditions, Strain Bb and P9 performed maximum current of (0.31 mA and 0.20 mA) and voltage of (0.46 and 0.29 V), respectively. Maximum cellulases activity (CMCase, FPase and β -glucosidase) were also observed in cultures grown under optimum conditions i.e. 1.7, 3.3 and 1.0 U/ μ l for Bb and 1.9, 4.2 and 1.5 U/ μ l for P9, respectively. Phylogenetic characterization of these isolates was carried out using 16S rRNA gene sequence. Results designated that Strain Bb belong to the genus *Bacillus* with the closest match with *Bacillus cereus* (97 % sequence homology). Strain P9 was found closely related to *Microbacterium resistens* (Strain AGP4-3) with 97 % sequence homology. Both *Bacillus cereus* and *Microbacterium resistens* were also capable of degrading cellulose and the latter was also able to produce electricity.

Keywords:

Microbial Fuel Cell, Cellulose-degrader, Electricity Generation, Phylogenetic Characterization.

1. Introduction

Large amounts of energy contained in different forms of biomass which can be referred to as waste usually remain unexploited and subject to microbial degradation. Direct exploitation of biological substrate degradation for electricity generation is a recent interest in Microbial Fuel Cells (MFC). Microbial Fuel Cell (MFC) is a device that converts biochemical energy into electrical energy with the aid of the catalytic reaction of microorganisms which provides new opportunities for the sustainable production of energy from biodegradable, reduced compounds that commonly present in wastes [1].

MFC consists of anode and cathode compartments separated by cation specific membrane. An anodic electrode potential is developed when the electrons from the oxidation of substrate by microorganisms are available to the electrode. These bacteria metabolize and transfer electrons to the anode that then pass through an external circuit to the cathode thus

producing current. Protons migrate through the solution to the cathode where they combine with oxygen and electrons to form water.

Electron cannot be transferred from the normal microbial electron transport systems to the electrode due to the non-conductive nature of the microbial cell surface. Electrochemical mediators were employed to render the electron transfer from the microbial cells to the anode [2]. The MFC that utilizes mediator as electron shuttle is called mediator-MFC. However, there are MFCs which mediator is excluded, known as mediatorless-MFC. In this type of fuel cells, an electrochemically-active microbe was used to directly transfer the electron by membrane-associated electron transfer [3].

Recently, most study has been focused on the generation of electricity by Fe (III)-reducing bacteria [1, 4], glucose and starch fermentating bacteria [2, 5], Sulfate-reducing bacteria [6] and others such as *Escherichia coli*, *Proteus vulgaris* and so on [2]. To the authors' knowledge up to now, the potential of cellulose degrading bacteria to produce electrical energy has not been reported elsewhere.

In this study, attempts were made to develop a MFC

* Corresponding Author. E-mail: adibah @ bio.fs.utm.my,
Tel: +607-5534157, Fax: +607-5566162

using two strains of cellulose degrader as the microorganisms, methylene blue as the mediator and various types of cellulosic compounds as the substrates. Further analysis by molecular techniques will be extensively studied to identify the relationship between both strains with others known electricity producing bacteria and cellulose degrader bacteria based on their 16S rRNA gene sequence.

2. Materials and Methods

2.1 Microorganisms

Locally isolated bacterial strains, Bb and P9 used in this study were capable of degrading cellulose. They were grown under facultatively anaerobic condition at 30 °C-37 °C. Growth and cellulose degradation study were carried out in Modified R2A liquid medium containing (g/L): Casamino acids (0.5), Yeast Extract (0.5), MgSO₄·7H₂O (0.05), K₂HPO₄ (0.3), KH₂PO₄ (0.09) and carboxymethylcellulose (5.0) as substrate.

2.2 Microbial Fuel Cell (MFC) System

The anode and cathode compartments (working volume of 75 ml each) were separated by cation-exchange membrane (Nafion®, Dupont Co., USA). Carbon electrodes (40 mm²) were used with copper wire connecting them through resistance and a multimeter. The cathode compartment contained a solution of 0.1 M Potassium ferricyanate in phosphate buffer (pH 7) as the electrolyte and the anode compartment contained 10 % (v/v) fresh bacterial inoculum in liquid medium. Methylene blue (0.18 % v/v) was added into the compartment as mediator. The MFC was placed in a temperature controlled chamber. Optimization of MFC was made by varying the microbial standard growth conditions as the factor of temperature, pH, carbon and nitrogen sources.

2.3 Instrumentation and Analysis

The potential between anode and cathode was measured using a multimeter (Proskit, China) and recorded every four hour intervals. An open circuit (OC) was used to measure potential (V) while current was measured in close circuit (CC) configuration. The MFC was discharged through an external resistance of 500 Ω to measure the current (I), in ampere unit (A).

2.4 Analytical Procedures

Cell-free supernatant of samples taken at regular intervals were used for the analysis of cellulases activity. Endoglucanase or CMCase activity was determined by measuring reducing sugar liberated from 2 % (v/v) CMC while FPase activity was determined by estimating the reducing sugar liberated from Whatman filter paper [7]. Both reactions were performed in 0.05 M sodium acetate buffer at pH 5 and incubated at 50 °C for 30 min (CMCase) or 1 hour (FPase).

The reducing sugar content was detected by using the 3,5-dinitrosalicylic acid method (DNS) with glucose as standard. One unit of CMCase and FPase was defined as 1 μmol reducing sugar released/ ml enzyme/ min.

β-glucosidase was determined by measuring the release of p-nitrophenol from p-nitrophenyl-β-D- glucoside [7]. One unit of β-glucosidase activity was determined as 1 μmol of p-nitrophenol liberated/ ml enzyme/ min.

Cellulose concentration was determined by two steps including by treating the samples with SDS (20 % v/v) to remove followed by removal of soluble sugars in the supernatant prior to quantify the reducing sugar in cellulose using phenol-sulphuric acid method with glucose as standard. Equivalent anhydroglucose was used for calculation [8].

2.5. Morphological characterization

A light microscope (Leica, Germany) was used to record morphological and behavioural characteristics of bacterial cells. Gram reaction and spore formation was observed according to standard methods. For scanning electron microscopy, an enriched electrode from MFCs was used. The specimen surface was dried in a dessicator. The micrographs were taken by using Smart Scanning Electron Microscope (Gemini Oxford Instrument, Germany) under magnification of x5 000 to x10 000.

2.6. 16S rRNA gene sequencing and analysis

Genomic DNA isolation was performed using DNA isolation kit following the manufacturer procedures (Promega, USA). The extracted genomic DNA of Strain Bb and P9 were used for PCR mediated amplification of 16S rDNA. The universal primers used were 5'-GAGTTTGATCCTGGCTCAG-3' (forward primers) and 5'-CGGCTACCTTGTTACGACTT-3' (reverse primers). The PCR products were confirmed with agarose gel electrophoresis and then sent for sequencing (1st Base, Malaysia) for partial 16S rDNA sequence.

The nested PCR primer was designed based on the results alignment of the partial sequences in order to amplify the middle region of 16S rDNA sequence of both strains. Self-design forward and reverse primers used for Strain Bb and P9 were 5'-GGAAACTGGGAGACTTGAG-3' (B647F) pair with 5'-GTTGCCATCATTAGTTG-3' (B867R) and 5'-ATTCCTGGTGTAGCGGTGG-3' (P652F) pair with 5'-GATTCGCTCCACCTCAC-3' (P1241R), respectively.

The sequences were compared with other known sequences deposited in GenBank databases using the basic local alignment search tool (BLAST). Similarity and distance matrices were calculated with the Clustal X (version 1.81). Phylogenetic tree was constructed based on the 16S rRNA gene sequences for electricity-producing bacteria and cellulose-degrading bacteria.

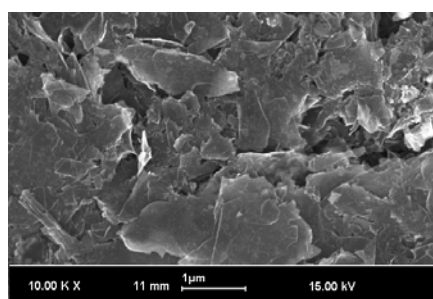
3.0. Results and Discussion

3.1 Characterization of Bacteria

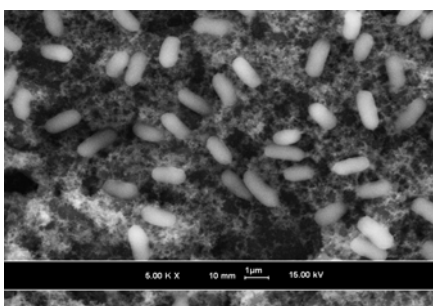
Strain Bb and P9 have been as the most potential cellulose degraders for electricity generation during metabolism. Both strains showed clearing zone when grown on the CMC-containing solid medium with Carmoisine red dye (0.5 % w/v) indicated to the production of cellulose. Morphological characteristic of Strain Bb and P9 showed

that they were distinguished by their colony-forming behaviour. Strain Bb appeared as creamy, round colonies with scalloped-like margin and umbonated surface. Strain P9 grown as smooth, round colonies with yellow pigment and have surfaces resembling water droplets. Cellular morphology analysis found that Strain Bb was a gram positive, rod shaped bacilli while Strain P9 was a gram negative cocci. Spores observed in Strain Bb were central endospores.

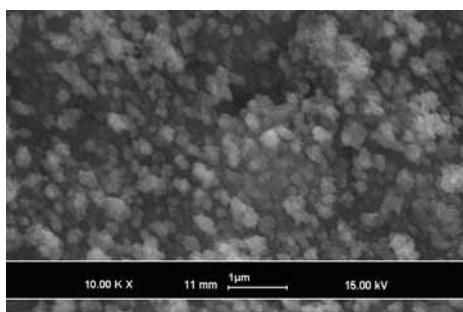
Growth of both strains as biofilm was observed in separate MFCs after two days of incubation. Formation of biofilm on the surface of carbon electrodes from anodic compartment were examined using a scanning electron microscope. The micrograph (Figure 1) obtained clearly confirmed the presence of microbial biofilm composed of pure culture of Strain Bb (Figure 1B) and Strain P9 (Figure 1C). An electrode from uninoculated MFC was used as control (Figure 1A) to confirm that this study was not affected by the presence of contaminant.



(A)



(B)



(C)

Figure 1: Electron micrographs of the electrode surface and particles separated from the electrode. The control electrode (A) and scanning electron micrographs of electrode and cells retrieved from the electrode; Strain Bb (B) and Strain P9 (C), respectively. Magnification used was $\times 10\,000$ (A, C) and $\times 5\,000$ (B).

Characterization based on 16S rRNA gene sequence was the most accurate method for microbial classification reported in many literatures [9]. The relationship of microorganisms could be summarized and concluded in a phylogenetic tree as showed in Figure 2. BLAST analysis of the 16S rRNA gene sequence of Strains Bb and P9 showed that they have the closest relations with *Bacillus cereus* and *Microbacterium resistens* with 97 % sequence homology, respectively. None of the literatures reported the ability of both bacteria to generate electricity during cellulose degradation.

Phylogenetic characterization of Strains Bb and P9 was carried out based on the phylogenetic tree constructed using the 16S rRNA gene sequences of unknown strains (Bb and P9) together with other known bacteria capable of electricity generation and/ or cellulose degradation (Figure 2). Construction of phylogenetic tree was carried out with *Geobacter sulfurreducens* as an outgroup. It has been widely reported that *Geobacter* species was commonly associated with electrode community in mediatorless-MFC and greatly distinguished with those of bacteria associated with mediator-containing MFC system [4].

From the phylogenetic tree, it was clearly showed that Strain Bb and P9 were belonged to Firmicutes phylum. Strain Bb was clustered together with other low G+C content *Clostridia* and *Bacilli* species including *Bacillus cereus* [10]. On the other hand, this group also exhibited similar cellular morphology of being gram positive, capable of producing endospores and grown under facultatively anaerobic condition [9]. In addition, Strain Bb also has similar ability to *Clostridium thermocellum* being capable of fermenting cellulose to produce solvent particularly ethanol [11].

In contrast, Strain P9 was clustered together with other high G+C content bacteria [10]. However it was closely branched from the group of *Microbacterium* sp. The genus *Microbacterium* was also previously known as *Aureobacterium* and they were phylogenetically intermixed [12]. On the other hand, *Streptomyces drozdowiczii* has been reported as a novel cellulolytic bacteria though none of the literature has mentioned its ability to generate electricity [13].

Based on several other characteristics of bacteria included in the phylogenetic tree such as metabolism, cellular morphology and G+C content characteristics, the relationship of Strain Bb and P9 could be justified and accepted.

3.2 Optimization of MFC

The performance of a MFC was considered in terms of microbiology and electrochemistry aspects. The electrochemical activity in MFC was optimized as the biological factors include temperature, pH, carbon sources and nitrogen sources.

3.2.1. Effect of temperature and pH

Temperature and pH were the major environmental factors that affecting physiological activity of most prokaryotes. At their optimum temperature and pH, they perform biological activities at the maximum rate such as growth and metabolism. Strain Bb and P9 was able to show good performance on electricity generation of temperature

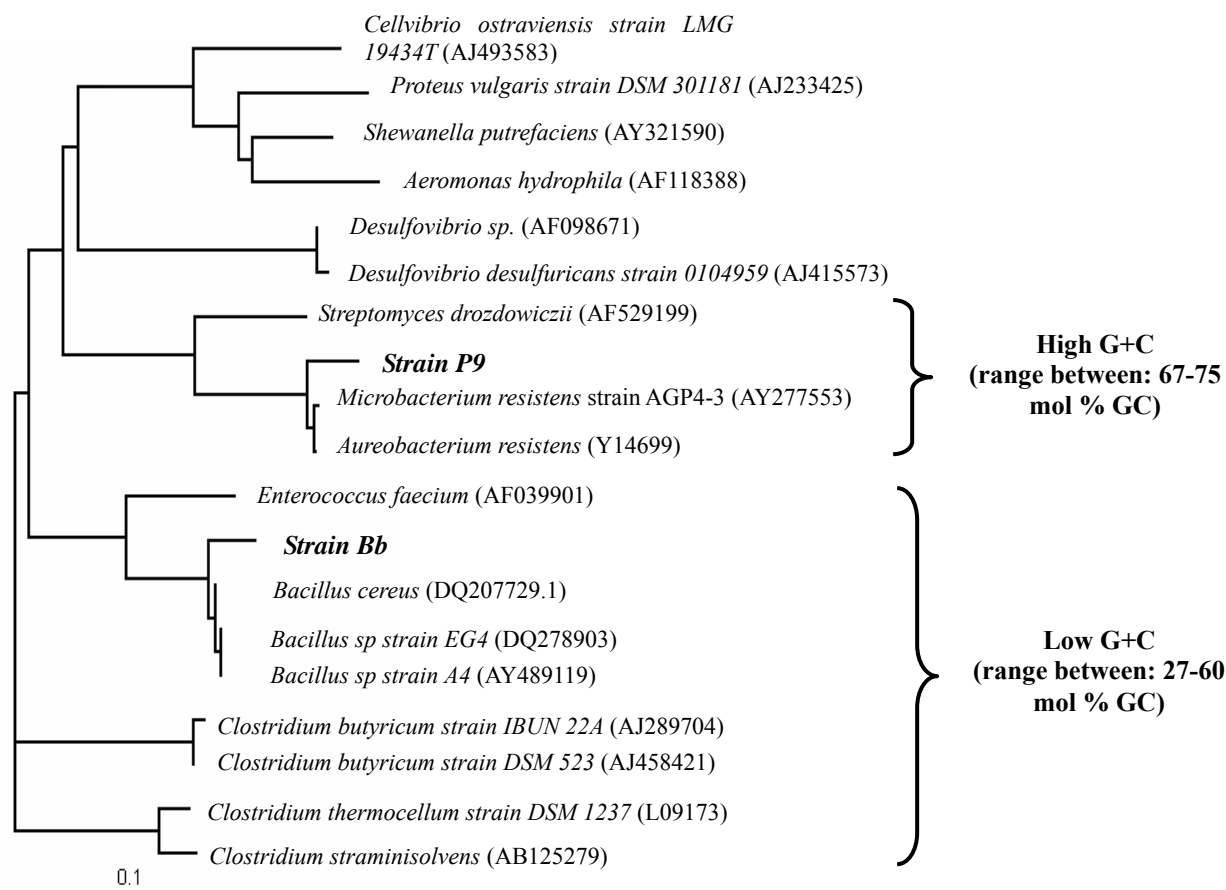


Figure 2: Phylogenetic distance tree based on the comparative analysis of 16S rRNA sequences of unknown strains (Bb and P9) together with other known bacteria capable of electricity generation and/ or cellulose degradation. Similarity and distance matrices were calculated with Clustal X (version 1.81). The tree was rooted with reference to *Geobacter sulfurreducens*. The scale bar represents the expected number of changes per sequence position.

ranging from 30 °C to 45 °C and the highest yield was obtained at 37 °C. Table 1 showed the relationship between cellulose metabolism and electrical energy produced by both strains Bb and P9, respectively. The highest cellulases activity was obtained at the optimum conditions. Bacterial Strain Bb and P9 have the same optimum temperature of 37 °C with the efficiency of electron transfer of 6.47×10^{19} and 6.29×10^{19} , respectively. At this temperature, acceptable value of potential and current were obtained from cultures of Bb (0.25 V, 0.12 mA) and P9 (0.18 V, 0.10 mA), respectively.

However, they were distinguished by their different optimum pH (Bb = pH 6.0 and P9 = pH 7.0) for growth and cellulose metabolism. Enhanced maximum electricity generation was observed in Strain Bb (0.35 V/ 0.13 mA) and P9 (0.20 V/ 0.12 mA). Maximum cellulases activity (CMCase, FPase and β -glucosidase) obtained were Bb = 1.7, 1.4 and 2.3 U/ μ l and P9 = 1.5, 1.4 and 1.8 U/ μ l. A total of 8.18×10^{19} (Bb) and 7.55×10^{19} electron were transfer through the circuit when only 31.5 % (Bb) and 33.3 % (P9) cellulose were degraded during fermentation.

Table 1: Comparison of voltage production (---●---) and cellulases, CMCase (---●---), FPase (---▲---) and β -glucosidase (---■---) activities (mU/ml) incubated at 37 °C.

T _{Initial} (°C)	Strain Bb				
	V max (V)	CMC ase (U/ μ l)	FPase (U/ μ l)	B-gluco sidase (U/ μ l)	
30	0.14	1.0	0.8	1.3	
37	0.25	1.4	1.1	1.8	
45	0.12	0.5	0.6	1.3	
50	0.10	0.5	0.6	0.8	
60	0.10	0.2	0.3	0.3	
	Strain P9				
	30	0.12	1.0	0.8	1.0
	37	0.18	1.2	0.8	2.0
	45	0.11	0.5	0.5	1.0
	50	0.07	0.5	0.5	0.8
	60	0.07	0.3	0.3	0.5

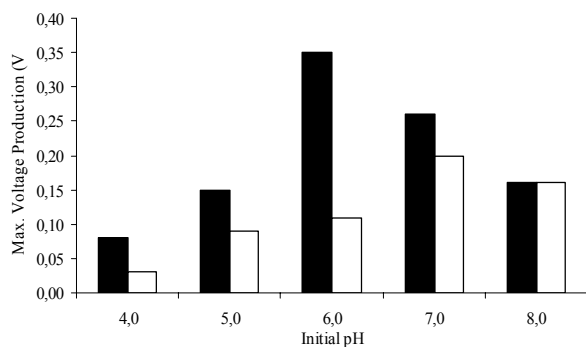


Figure 3: Effect of initial pH on voltage production by Bb (■) and P9 (□) grown at 37 °C.

From Figure 3, it was clearly demonstrated that, at pH 5.0 and below, electrochemical and cellulases activity were found lower compared to other initial pH tested. This may be due to the denaturation of cellulases protein or active site(s) under acidic condition [14]. According to the results obtained, electrochemical activity could be attributed by the enzymatic activities contributed by both strains. At optimum temperature and pH, cellulose was in its most optimum conformation for display maximum activities leading to the highest metabolic rate by the microbes and in net electron production.

3.2.2. Effect of carbon sources

Two types of cellulose with different degree of crystallinity were used which were substituted cellulose (CMC) and microcrystalline cellulose (ethyl cellulose, native cellulose and empty fruit bunch). The electrochemical activity of different types of cellulose used was summarized in Table 2. Electricity production was found almost two times higher in native cellulose (Bb) and

Table 2: Comparison of voltage and cellulases production by strains Bb and P9 in different types of carbon sources grown at optimum temperature and pH.

Types of Carbon Sources	Strain Bb				
	V max (V)	CMCase (U/μl)	FPase (U/μl)	B-glucosidase (U/μl)	
CMC	0.32	1.7	1.4	3.0	
Ethyl Cellulose	0.26	1.5	1.9	3.3	
Native Cellulose	0.44	1.7	3.3	4.8	
Cellulose EFB	0.26	1.7	3.3	5.0	
Types of Carbon Sources	Strain P9				
	V max (V)	CMCase (U/μl)	FPase (U/μl)	B-glucosidase (U/μl)	
	CMC	0.21	1.4	1.4	2.8
	Ethyl Cellulose	0.21	1.2	2.2	4.0
	Native Cellulose	0.28	1.7	3.9	4.5
	Cellulose EFB	0.28	1.9	3.9	5.8

empty fruit bunch (P9) compared to CMC. It also showed an increased of cellulases activity i.e. 1.7, 3.3 and 4.8 U/μl (Bb) and 1.9, 3.9 and 5.8 U/μl (P9) for CMCase, FPase and β-glucosidase, respectively when grown in the medium supplemented with their optimum temperature, pH and carbon source.

This result implied that, cellulases and electrochemical activities were strongly depended on the type of substrate supplemented in the medium. It was demonstrated that the high crystallinity of cellulose (native cellulose and EFB) would actively induced cellulases activity. The stoichiometry of cellulose is commonly described by the following net reaction: $[-C_6H_{11}O_6-]_n + a H_2O \rightarrow b CH_3COOH + CH_3CH_2OH + d CO_2 + H_2 + (e^-)_n$ [15]. Therefore cellulose metabolism and cellulases activity was accompanied in order to increase electrochemical activity as an electrons donor by both strains.

3.2.3. Effect of nitrogen sources

Two types of nitrogen sources were investigated i.e. casamino acids, peptone and yeast extract (represents organic nitrogen sources) and NH₄Cl and (NH₄)₂SO₄ (represents inorganic nitrogen sources). In addition, combinations of organic nitrogen sources (casamino acids plus peptone, yeast extract plus peptone and yeast extract plus casamino acids) were also studied. Based on the results obtained (Figure 4), combinations of casamino acids and peptone (Bb = 0.46 V/0.31 mA) and yeast extract and peptone (P9 = 0.29 V/0.20 mA) enhanced the maximum electricity generation by the bacteria. The used of combinations of nitrogen sources were also described in literature for enhanced growth and metabolism of microorganisms [9].

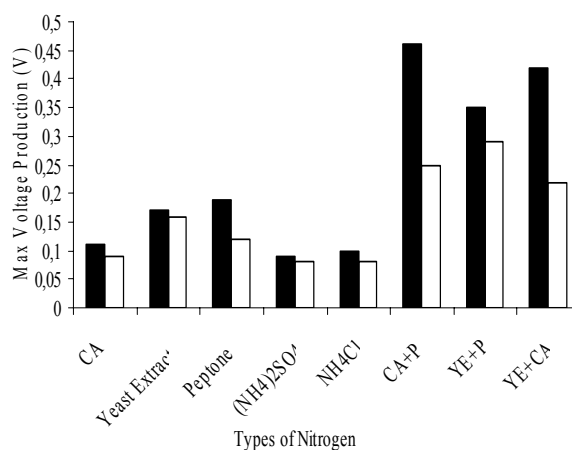


Figure 4: Comparison of maximum voltage production using different types of nitrogen sources by Bb (■) and P9 (□) grown under optimized conditions.

Under these conditions, Strain Bb and P9 sustained a 1.95×10^{20} and 1.44×10^{20} electron through the circuit. This was followed by the increased in cellulases (CMCase, FPase and β-glucosidase) activity i.e. Bb (1.7, 3.3 and 6.0 U/μl) and P9 (2.0, 3.9 and 5.8 U/μl). A combination of two types of

organic nitrogen increased electrochemical activity of the bacteria by approximately two fold compared to those supplemented with single organic nitrogen. These results suggested that various organic compounds in nitrogen sources provided an extra carbon source, vitamins and mineral required by the bacteria for growth and cellulose metabolism. In addition, combination of casamino acid + peptone that preferred by Strain Bb provided high content of free amino acids were important for protein formation by the bacteria. On the other hand, mixture of yeast extract and peptone will provide the bacteria with high vitamin B that important as growth factor.

4.0. Conclusion

This study has successfully carried out to name and locate both Strains Bb and P9 appropriately in a phylogenetic tree. Based on the 16S rRNA gene analysis, it is possible to confirm that Strain Bb and P9 are *Bacillus cereus* and *Microbacterium resistens* with the highest sequence homology of 97 %, respectively. However, further characterization of both strains will possibly enable the researchers to establish their novelties and to propose these strains as novel species of *Bacillus* and *Microbacterium*, respectively. Electricity generation by Strains Bb and P9 were increased by 77 % and 71 % respectively under optimum conditions. It was also found that metabolism of cellulose by both strains was directly related with the generation of electricity. MFC could be functioned on different carbohydrates and also on complex substrates present in wastes. Depending on the operational parameters of the MFC, different metabolic pathways were used by the specific organisms. The MFC technology was evaluated relative to current alternative for energy generation.

Acknowledgments

The author wish to thank Dr. Adibah Yahya and Assoc. Prof. Dr. Zaharah Ibrahim for their strong support, useful comment and critically reading this manuscript. Thanks are also due to Dr. Astuty Amrin and Mr. Jefri, Faculty of Mechanical Engineering for their assistance in acquiring SEM data.

References

- [1] Kim, H. J., Park, H. S., Hyun M. S., Chang, I. S., Kim, M. and Kim, B. H. 2002. A Mediatorless-Microbial Fuel Cell Using a Metal-reducing Bacterium, *Shewanella putrefaciens*. *Enzyme and Microbial Technology*. 30: 145-152.
- [2] Katz, E., Shipway, A. N. and Willner, I. 2003. *Handbook of Fuel Cells-Fundamentals, Technology and Applications*. John Wiley and Sons Ltd.
- [3] Rabaey, K. and Verstraete, W. 2005. Microbial Fuel Cells: Novel Biotechnology for Energy Generation. *TRENDS in Biotechnology*. 23 (6): 291-298.
- [4] Bond, D. R. and Lovley, D. R. 2003. Electricity Production by *Geobacter sulfurreducens* Attached to Electrodes. *Applied and Environmental Microbiology*. 69 (3): 1548-1555.
- [5] Niessen, J., Schroder, U. and Scholz, F. 2004. Exploiting Complex Carbohydrates for Microbial Electricity Generation – A Bacterial Fuel Cell Operating on Starch. *Electrochemistry Communications*. 6: 955-958.
- [6] Cooney, M. J., Roschi, E., Marison, I. W., Comninellis, Ch. and Urs von Stockar. 1996. Physiologic Studies with the Sulphate-reducing Bacterium *Desulfovibrio desulfuricans*: Evaluation for Use in a Biofuel Cell. *Enzyme and Microbial Technology*. 18: 358-365.
- [7] Umikalsom, M. S., Ariff, A. B., Zulkifli, H. S., Tong, C. C., Hassan, M. A., Karim, M. I. A. 1997. The Treatment of Oil Palm Empty Fruit Bunch Fibre for Subsequent Use as Substrate for Cellulase Production by *Chaetomium globosum* Kunze. *Bioresource Technology*. 62: 1-9.
- [8] Desvaux, M., Guedon, E., Petitdemange, H. 2000. Cellulose Catabolism by *Clostridium cellulolyticum* Growing in Batch Culture on Defined Medium. *Applied and Environmental Microbiology*. 66 (6): 2461-2470.
- [9] Madigan, M. T., Martinko, J. M, Parker, J. 9th. Ed. 2000. *Brock Biology of Microorganisms*: Southern Illinois.
- [10] Funke, G., Graevenitz, A. V., Clarridge III, J. E. and Bernard, K. A. 1997. Clinical Microbiology of Coryneform Bacteria. *Clinical Microbiology Reviews*. 10 (1): 125-159.
- [11] Weimer, P. J. and Zeikus, J. G. 1977. Fermentation of Cellulose and Cellobiose by *Clostridium thermocellum* in the Absence and Presence of *Methanobacterium thermoautotrophicum*. *Applied and Environmental Microbiology*. 33 (2): 289-297.
- [12] Takeuchi, M. and Hatano, K. 1998. Union of the Genera *Microbacterium* Orla-Jensen and *Aureobacterium* Collins *et al.* in a Redefined Genus *Microbacterium*. *International Journal of Systematic Bacteriology*. 48: 739-747.
- [13] Lima, A. L. G. D., Nascimento, R. P. D., Bon, E. P. D. S. and Coelho, R. R. R. 2005. *Streptomyces drozdowiczii* Cellulose Production Using Agro-industrial By-products and Its Potential Use in the Detergent and Textile Industries. *Enzyme and Microbial Technology*. 37: 272-277.
- [14] Hu, Z. H., Yu, H. Q., Zhu, R. F. 2005. Influence of Particle Size and pH on Anaerobic Degradation of Cellulose by Ruminant Microbes. *International Biodeterioration and Biodegradation*. 55: 233-238.
- [15] Angenet, L. T., Karim, K., Al-Dahhan, M. H., Wrenn, B. A. and Domiguez-Espinosa, R. 2004. Production of Bioenergy and Biochemicals from Industrial and Agricultural Wastewater. *TRENDS in Biotechnology*. 22 (9): 477-485.

Contribution from the Domestic Sector Considering various Scenarios and by using effective tool (SAP Energy Rating) to meet the Kyoto Commitment

Dr. S. R. Samo¹, Abdullah Saand²

¹Department of Energy and Environment Engineering, QUEST, Nawabshah Sindh, Pakistan
Tel 0092 244 9370361, cell 0333 7010738 fax 0092 244 9370367

Email: srsamo@yahoo.com

²Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

Email: abdullah.s@siswa.utm.my

Abstract

There has been much debate over the scope for reducing emissions from the domestic sector. In Pakistan the domestic sector accounts for more than 20% of total energy consumption [1] and about 30% in the UK. The UK government has signed the Kyoto Protocol to the Rio convention and it also has a commitment with the aim of reducing CO₂ emissions to 20% below the 1990 levels by 2010 [2]. As buildings are responsible for approximately half of the UK CO₂ emissions, of which the domestic sector accounts for 50%, meeting the above mentioned target is likely to require a significant contribution from the building sector. In this study different scenarios were each examined to assess the possible contribution of the domestic sector to achieving the UK government's target.

This paper gives an analysis of various energy efficiency measures to achieve the contribution of the domestic sector to the UK government's target. Improvements of the wall U-values of the older dwellings can give more than 20% reduction in energy consumption and CO₂ emissions in the total domestic sector. By considering one of the scenarios, the target to reduce the CO₂ emission can also be achieved well before the year 2010. In this paper the impact of the thermal insulation, and efficiency & control of the heating system on SAP energy rating have also been investigated. This study has also pointed out some areas where the SAP energy rating could be improved, for example the fabric thermal insulation could increase the SAP rating from 33 to 80.

Keywords: Kyoto, Energy, CO₂ Emission, Building Regulations, Domestic Sector

1. Introduction

A study was conducted to analyse the effect of the UK thermal Building Regulations on national energy consumption. This paper presents the percentage of dwelling stock, energy consumption, and carbon dioxide emission in different types of dwellings, which satisfy different Building Regulations. Alone among the major European nations, the UK is more than self-sufficient in energy supply. Yet energy demand in the UK has fallen, by 1% a year on average [3]. There have been several changes in the Building Regulations, which have more than doubled, and even tripled, thermal standards over the last 30 years. Table 1 illustrates the comparison between the UK's first Building Regulations 1965 [4] and the current (1995) Building Regulations [5]. However these apply only to new buildings, leaving the vast majority of the building stock (about 66%) untouched.

For this research work a computer model, "Housing Energy Consumption Analyser (HECA)" was produced to calculate the above-mentioned parameters, while considering the requirement of UK Building Regulations from 1965 to 1995. U-value improvement could save energy and reduce CO₂ emission in dwellings [6]. The effect of decreased U-values of pre-1965 dwellings to achieve the government target was analysed and results are presented in this paper. In this study different scenarios were also each examined, considering the various methodologies, to reduce the carbon dioxide emission in domestic sector.

2. Methodology

The published housing data of total number of dwellings [7] from 1985 to 1998 was used to calculate the total energy consumption and CO₂ emission of UK dwellings. Four revisions of Building Regulations - Building Regulations 1965 [4], Building Regulations 1976 [8], Building Regulations 1985 [9] and Building Regulations 1995 (as amended) [5] have been introduced in this country. Therefore, the age and thermal performance distribution of the dwelling stock was classified according to their compliance with the Building Regulations.

Six types of dwellings - Detached house, Semi-detached house, Bungalow, Flat, Post-1919 Terrace house and Pre-1919 Terrace house - were considered. The construction details and the U-values of the construction elements according to the different building regulations are summarised in Table 2. The percentage distributions of different types of dwellings were considered to be the same as described in [10] to calculate the total number of dwellings by type. The seasonal energy consumption for a single dwelling of each type was calculated by the method described in [11]. The CO₂ emission was calculated by using the Elmhurst SAP energy rating computer program [12]. The total energy consumption of all UK dwellings was calculated by adding the energy consumption of each part of the dwelling stock satisfying different Building Regulations from 1965 to 1995 and the energy

consumption of the pre-1965 dwellings as shown in Eq.(1), i.e.,

$$EC(\text{total}) = \Sigma (EC_{95} + EC_{85} + EC_{76} + EC_{65} + EC_{\text{pre-65}}) \quad (1)$$

$$EC_{95} = \Sigma (EC_{\text{det}95} \times DS_{\text{det}95} + EC_{\text{semidet}95} \times DS_{\text{semidet}95} + EC_{\text{bungalow}95} \times DS_{\text{bungalow}95} + EC_{\text{flat}95} \times DS_{\text{flat}95} + EC_{\text{post 1919 terrace}95} \times DS_{\text{post 1919 terrace}95} + EC_{\text{pre 1919 terrace}95} \times DS_{\text{pre 1919 terrace}95})$$

Where

EC_{95} = Energy consumption of dwelling stock satisfying the Building Regulations 1995

$EC_{\text{det}95}$ = Energy consumption of detached house satisfying Building Regulations 1995

$DS_{\text{det}95}$ = The number of detached houses satisfying Building Regulations 1995

The following three scenarios were each examined to assess the possible contribution of the domestic sector to achieving the UK government's target.

- 1) To demolish the old dwelling (pre-1965) and to construct new dwellings satisfying the Building Regulations 1995 at the rate of 180,000 dwellings a year
- 2) To improve the U-values of construction elements of pre-1965 dwellings to satisfy the Building Regulations 1995 at the rate of 400,000 dwellings a year
- 3) To improve the U-values of construction elements of pre-1965 dwellings to satisfy the Building Regulations 1995 at the rate of 200,000 dwellings a year and gradually replace the pre-1965 dwellings with new construction of dwellings according to the Building Regulations 1995 at the rate of 50,000 dwellings a year.

3. Results and Discussion

In the UK, four revisions of the national building regulations have been introduced since 1965 at intervals of about 10 years. Each successive revision has improved the U-values of construction elements to reduce the energy use and its related impact on the environment. The improvement in U-values of construction elements of dwellings by each successive revision of the building regulations is shown in Fig.1. Since 1965 the roof U-values have been decreased by about 86% and the U-values of walls and floor are decreased by about 75% each, whereas the U-value of windows have been decreased only by 46% by changing single-glazed windows to double-glazed windows. As the successive revisions of the UK building regulations have reduced the U-values of the external envelope, it is well known that the typical ventilation conductance has reduced also. This paper concentrates on the fabric conductance, as the ventilation performance is far less quantified than the fabric

performance. The effect of modulating the U-values of construction elements on energy consumption and CO₂ emission in the UK dwellings was analysed.

The effect of decreased U-values of construction elements of different types of dwellings to satisfy different building regulations on the energy consumption and carbon dioxide emission in each type of dwelling was also analysed. It was observed that a detached house satisfying the building regulations 1965 consumes about 80 GJ/year of energy for heating and produces about 8 tonnes of carbon dioxide. When the same example house was modified to give U-values which satisfy the requirement of the current building regulations (1995), these figures were both reduced by about 65%. The example Bungalow consumes the highest energy per unit of floor area and the example Detached house consumes the lowest as compared to other type of dwellings. The energy consumption in Pre-1919 terrace house was -about 10% higher than the energy consumption in the Bungalow and about 12% less than the energy consumption in a detached house. If the pre-1919 terrace house could be upgraded to reach the standard of the 1995 regulations then the reduction in total carbon dioxide emission would be about 60%.

The Energy consumption and carbon dioxide emission of the dwellings in Great Britain was calculated using a computer program. The total number of dwellings of Great Britain in 1995 was about 24 millions [13]. The annual energy consumption of these dwellings is more than 1500 PJ - costing about £14 billions per year and producing about 170 million tonnes of carbon dioxide annually. The percentage distribution of existing dwelling stocks satisfying different Building Regulations from 1965 to 1995 is shown in Figure 2. About 66% (16 million) of total dwellings are old dwellings (pre-1965). Only 15% (3.7 million) of total dwellings satisfy the UK's first Building Regulations 1965. More than 9% (2 million) dwellings satisfy the Building Regulations 1976. Whereas less than one million (4%) are constructed to satisfy the current Building Regulations 1995 (as amended). The pre-1965 dwellings account for more than 70% of total energy consumption. The dwelling stock, which satisfies the Building Regulations 1965, consumed about 17% (about 250 PJ/Yr) of total energy consumption. Ten percent (about 160 PJ/Yr) of total energy consumption was accounted for by the remaining three dwelling stocks satisfying the Building Regulations 1976, 1985 and 1995. The carbon dioxide emission in the above mentioned dwelling stocks were also analysed and results are shown in Figure 2.

The dwellings, which satisfy the Building Regulations 1995, produced only less than 3 million tonnes of carbon dioxide emission per year (1.5% of total). Whereas the old dwellings (pre-1965) were responsible for about 130 million tonnes of Carbon dioxide emission per year (about 75%). The dwelling stock, which satisfies the Building Regulations 1965, produced about 15% (25 million tonnes/Yr.) of total carbon dioxide emission. The energy consumption and carbon dioxide emission of the dwelling stock, which satisfy the Building Regulations 1976, was about 7% respectively. About half of the remaining

portion of dwelling stock i.e., 6% (about 1.4 million dwellings) satisfy the Building Regulations 1985 and other half i.e., 4% satisfy the Building Regulations 1995 (as amended). This remaining portion of the dwelling stock (10%) consumes about 4% (60 PJ/Yr.) of total energy used and is responsible for about 4% (7 million Tonnes/Yr.) of total carbon dioxide emission.

The effect of individually decreased thermal insulation of walls, floor and roof (windows were considered either single or double-glazed) on the energy consumption in the pre-1965 dwelling stock was analysed by decreasing each of the U-values by between 20 and 80%. Figure 3 shows the effect of individually decreased U-values on the energy consumption of pre-1965 dwelling stock.

By decreasing the U-values of walls by 20%, the energy consumption of this old dwelling stock can only be decreased by 12%. Forty percent decrease in U-values of this construction element will reduce the energy consumption by up to 20%. Quite significant savings of the energy (about 40%) in this dwelling stock could be made if the U-values of walls can be decreased by 80%. If the U-values of roof and floor could be decreased by 80%, then up to 13% savings of energy consumption in the pre-1965 dwelling stock can be made (Fig.3). By considering double glazed windows instead of single glazed in pre-1965 dwelling stock the energy consumption decreased by only about 9%.

The effect of U-values of various construction elements of pre-1965 dwelling stock on the reduction of CO₂ emission in this particular stock was also analysed. The decrease in U-values of walls has the highest impact on reduction of energy consumption as compared to the other construction elements, which is to be expected, because of the larger area of this element. If the U-values of walls could be decreased by 40% then the energy consumption in this dwelling stock can be reduced by 20%. As a result a reduction in total CO₂ emission in pre-1965 dwelling stock of about 30% and in overall domestic sector of about 23% can be made as shown in Fig.4. This energy efficiency measure will be enough to achieve the contribution of the domestic sector to the government's target. The scenario 2 was analysed and results are shown in Fig.5, which shows the targeted and predicted CO₂ emission. The vertical dotted line shows the year 2010 to achieve the target. It can be seen from Fig.5 that by considering this scenario the target to reduce the CO₂ emission can be achieved well before the year 2010.

4. Conclusions

Following conclusions were made:

- Building Regulations apply only to new buildings, leaving the vast majority of the building stock (about 66%) untouched.
- The old dwellings, which were constructed before 1965, consume 73% of total heating energy in housing and are responsible for about 75% of carbon dioxide emission.

- Only 4% of the dwelling stock satisfy the Latest Building Regulations 1995, consume 2% of total energy consumption, and produce about a similar percentage of carbon dioxide emission in housing.
- Since 1965 the roof U-values have been decreased by, about 86% and the U-values of walls and floor are decreased by about 75% each. The U-values of windows have been decreased only by 46% by changing single-glazed windows to double-glazed.
- A typical standard detached house satisfying the building regulations 1965 produces about 8 tonnes of carbon dioxide. The same example house satisfying the requirement of the current building regulations (1995) produces only about 3 tonnes of carbon dioxide.
- If the U-values of walls could be decreased by 40% then the energy consumption in this dwelling stock can be reduced by 20%. As a result a reduction in total CO₂ emission in pre-1965 dwelling stock of about 30% and in overall domestic sector of about 23% can be made This energy efficiency measure will be enough to achieve the contribution of the domestic sector to the governments target. The scenario 2, examined in this study, seems to be the reasonable method to achieve the contribution towards the government's target of reducing the CO₂ emission level to 1990 level by the year 2010.

References

- [1] "Pakistan Energy Yearbook 2004, Hydrocarbon Development Institute of Pakistan, Islamabad.
- [2] HMSO, UK Climate Change Programme, Consultation Paper, Department of the Environment, Transport and the Regions, St Clements House, 1-16 Colegate, Norwich, UK, 1998
- [3] J. O. Lewis and J. Gouiding, European Directory of Energy Efficient Building 1993, James & James, London, UK 1993
- [4] C Kennard & J. Dufton, The Building Regulations, Butterworths, London, 1966.
- [5] HMSO, The Building Regulations 1991, Conservation of Fuel & Power, Approved Document Part L1, 1995 Edition, UK, 1995.
- [6] Letherrnan KM and Samo S.R., The Effect of the UK Building Regulations on Environmental Change, paper presented at the workshop of GIB W67, Czech Technical University, Prague, Czech Republic, 1998.
- [7] HMSO, Housing and Construction Statistics 1985-1995 Great Britain, Department of the Environment, St. Clements House, 2-16 Colegate, Norwich, UK, 1997.
- [8] A J Elder, Guide to the Building Regulations 1976, The Architectural Press: London, 1979.

- [9] HMSO, The Building Regulations 1985, Conservation of Fuel & Power, Approved Document Part L1, 1991.
- [10] E A Alien and A A Pinney, Standard Dwellings for Modeling: Details of Dimensions, Construction and Occupancy Schedules. Building Res. Establishment, Garston, Watford, UK, 1991.
- [11] BRE , Heat Losses From Dwellings, Building Research Establishment (BRE) Digest 190, Department of the Environment, Building Research Establishment, HMSO Benham and Company Limited, Colchester, UK. 1976.
- [12] Elmhurst Energy Systems Limited, SAP Energy Rating Software, version 1.08 (Design),, 1996, Elmhurst Energy Systems Limited, Elmhurst Farm, Bow Lane, Witherbrook, Nr. Coventry, UK, 1996
- [13] HMSO, English House Condition Survey 1991, Energy Report Department of the Environment, St. Clements House, 2-16 Colegate, Norwich, UK. 1996.

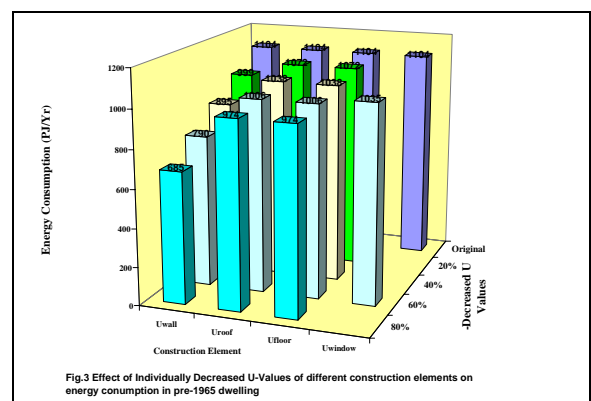
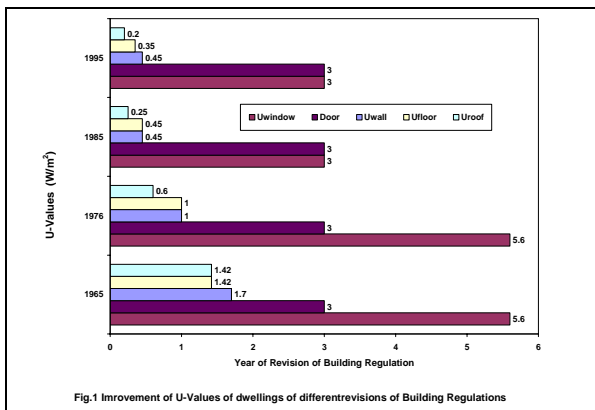
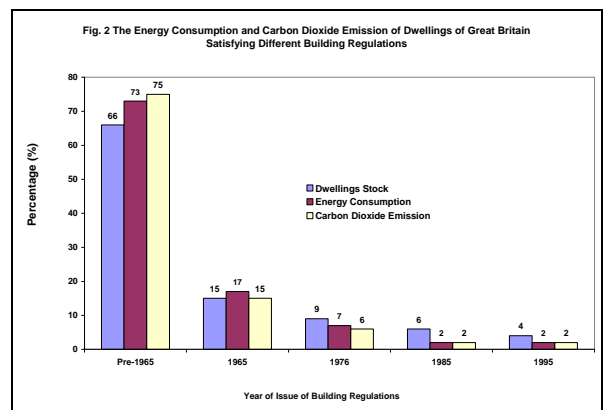
Table 2 Areas, Volumes and Thermal Conductance of Different Types of Standars Dwellings

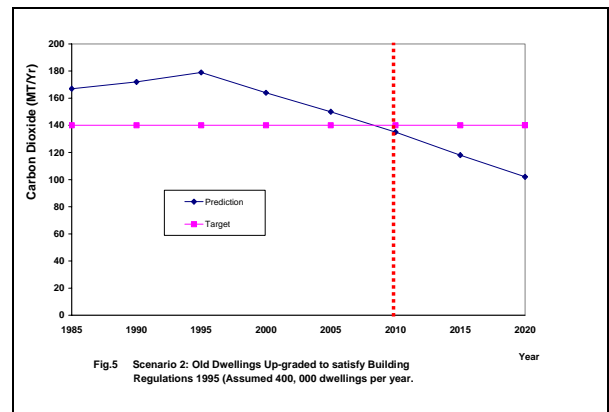
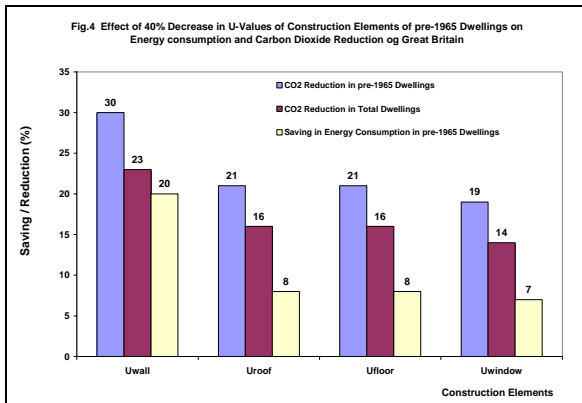
Dweeling Type	Total Floor Area (m ²)	Volume (m ³)	Wall to Floor ratio	Fabric Thermal Conduc tance (W/K)
Detached	99	240	1.4	433
Semi-Detached	85	199	1.6	378
Bungalow	63	149	1.2	345
Post 1919 Terrace	75	173	1.5	337
Pre 1919 Terrace	86	212	1.9	414

Source: BEPAC Research Report, 1991 [10]

Table 1 U-Values of construction elements according to different Building Regulations

Year of Building Regulation	U-Values (W/m ² K)				
	Wall	Window	Floor	Roof	Door
1965	1.7	5.6	1.42	1.42	3
1995	0.45	3	0.35	0.2	3





Influence of Chlorobenzene Levels on Bioremediation by Microbial Consortia from Residential Wastewater

Noor Hafiza Harun, Firdausi Razali*

Department of Bioprocess Engineering
Faculty of Chemical and Natural Resources Engineering
University of Technology Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5535513, Fax: +60-7-5581463, E-mail: fiezhar@yahoo.com

Abstract

The influence of chlorobenzene (CB) levels on microbial consortia originated from local residential wastewater on growth and degradation kinetics was investigated. The growth and degradation kinetics were monitored in batch cultures supplemented with CB as sole carbon source at concentrations of 0, 0.05, 5.0, 15, and 30% (v/v). The highest final cell concentration, namely 6.3 g/L, was at 30% (v/v) CB, while no significant growth was witnessed at concentrations of 0, and 0.05% (v/v). It is suggested that low CB concentration does not provide sufficient energy for growth. As well, an increase of CB concentration utilized represents the increment of the efficiency of CB biodegradation. This concluded that optimum degradation occurred around 30% (v/v). This also indicated that the consortia were robust enough to grow on this recalcitrant toxic aromatic compound without any prior adaptation. The outcomes of this study provide useful information and guideline in employing the consortia from residential wastewater to solve chlorobenzene contaminated in the environment.

Keywords:

Microbial consortia, chlorobenzene levels, residential wastewater

1. Introduction

Contamination of soils, groundwater, sediments, surface water and air with toxic halogenated aromatic compound become the major problems facing the industrialized world today. Therefore, the destruction of these pollutants was emphasized and executed under safety conditions in order to protect human and environment from the hazardous health effects.

One possible solution called bioremediation has become increasingly important rather than chemical and physical processes. This biological process was approached because of its perceived low cost, simplicity and its low adverse effect on the environment [1]. There are numerous applications of bioremediation treatment technologies including bioaugmentation, biofilters, biostimulation, bioreactors, bioventing, composting, and landfarming [2].

The aromatic compounds that are currently being targeted by bioremediation because of its resistances include the chlorinated benzenes [3], which are the priority environmental pollutants listed by the Environmental Protection Agency (EPA) [4]. The extensive used of

chlorobenzenes as organic solvents, insecticides, degreasers and deodorants, and their production as intermediates in the synthesis of chemicals such as dyes and pesticides, has led to a widespread release of these xenobiotic compounds into the environment [4,5].

These compounds have been found in a wide range of environmental media at high concentration including soils [6], groundwaters [7], sewage sludge [8,9], marine and lake sediments [10,11], and open water columns [12,13]. They are also known as potentially important riverine contaminants especially found in United Kingdom (UK) [14].

Thus, the responsibility of microorganisms for their removal from the environment via enzymatically reactions appears to be very essential. Native microorganisms can degrade CB as sole carbon and energy source in various substrates such as soil, sediment, sewage sludge, and water.

The major mechanism of aerobic CB degradation, which via oxidative dechlorination was usually initiated by dioxygenative hydroxylation, then leading to the formation of catechols. Finally, for completion of the CB biodegradation process, it undergoes the ring fission and subsequent mineralization to carbon dioxide and water. Besides that, the CB biodegradation under anaerobic condition also occurred but at a slower rate [15].

As a result, the microbial degradation of many chlorinated benzenes was examined and the results

* Corresponding Author. E-mail: r-firdaus@utm.my,
Tel: +60-7-5535513, Fax: +60-7-5581463

reported that different bacterial strains such as *Pseudomonas* sp., *Alcaligenes* sp., and *Xantobacter* sp. are individually able to use chlorobenzenes as growth substrates [16,17,18]. Moreover, the indigenous microbial communities especially from the CB contaminated sites, which acted as synergism teams, are also capable to degrade CB [19,20,21,22].

In this paper, the degradative capabilities of indigenous microbial consortia from residential wastewater in CB biodegradation under aerobic closed-system and batch culture conditions are reported. The objective of the study was to evaluate the significance of CB levels in bioremediation system.

2. Materials and Methods

2.1 Chemicals and Medium

The 99.9% purity CB with analytical grade purchased from Fischer Scientific, Germany was used in corresponding concentrations as sole source carbon throughout this study. For all cultivations, the liquid mineral medium utilized was sterilized and supplemented for enrichment. These medium consist of 3.3g/L di-kaliumhydrogenphosphat (K_2HPO_4), 1.9g/L natrium dihydrogenphosphat ($NaH_2PO_4 \cdot 2H_2O$), 4.5g/L ammonium sulfat ($(NH_4)_2SO_4$) and 0.2g/L magnesium-heptahydrat ($MgSO_4 \cdot 7H_2O$). Besides, yeast extract was added as an additional nitrogen source at concentration of 5.0 g/L.

2.2 Source of microorganisms

The source of microorganisms described in this study were from local residential wastewater, which were collected from pond of Desa Bakti, UTM. These samples content a mixed culture of miscellaneous unknown species. Therefore, the synergism work from these indigenous microbial consortia might be useful for degradation of toxic aromatic compounds in the environment. Apparently, the pollutants in wastewater may classified into five categories:

- a) organic matter (BOD),
- b) disease-causing microorganisms (pathogens),
- c) nutrients,
- d) toxic contaminants, and
- e) dissolved minerals.

2.3. Culture of microorganisms

2.3.1. Enrichment

All cultivations were enriched in a 500mL shake flasks in laboratory scale under batch conditions. The environmental conditions were controlled at room temperature, and pH between 6 to 8. Moreover, maintaining the aerobic circumstances during the experimental works are also emphasized in this study. This enrichment process was implemented for 24 hours to get the optimum microbial populations.

2.3.2. CB Biodegradation

Batch cultures was also employed in 500mL Erlenmeyer flasks for full treatment of CB biodegradation. The flasks containing 250mL working volume were inoculated with enriched inoculum, and supplemented with CB as sole carbon and energy source at concentration of 0, 0.05, 5.0, 15, and 30% (v/v). The experimental works were carried out on orbital shaker for 6 days. The environmental conditions were fixed as designed in enrichment process.

2.4 Measurement of growth

Growth of cultures in this study was monitored spectrophotometrically at wavelength of 600nm by Shimadzu UV-160 Spectrophotometer. The optical density (OD) is referred to standard curve of OD versus biomass concentration. As a rule of thumb, an OD of 1 unit corresponds to approximately 1 g/L of cell dry weight.

2.5 Measurement of CB concentration

CB concentration was calculated by detecting the absorption unit (AU) using High Pressure Liquid Chromatography (HPLC) at UV wavelength of 254nm [23]. Column from types of C18 with 3.9mm by 30cm was used and mobile phase for isocratic detection was methanol:water with ratio of 70:30. The flowrate was fixed at 1.0mL per minute with volume injection at 5.0 μ L during runtime for 5 minutes.

2.6. Measurement of kinetic growth and degradation

Various mathematical models have been proposed to quantitatively describe microbial growth kinetics. The Monod model (equation 1) is considered the basic equation [24], which has been improved by including expressions for e.g., maintenance, diffusion, or transport limitation [25]. This Monod equation is semiempirical; it is derived from the premise that a single enzyme system with Michaelis-Menten kinetics is responsible for uptake of the residual substrate concentration (s), and the amount of that enzyme or its catalytic activity is sufficiently low to be growth-rate limiting.

$$\mu_g = \frac{\mu_{max} s}{K_s + s} \quad (1)$$

Whereas, μ_g is the specific growth rate in $hour^{-1}$, μ_{max} is the maximum specific growth rate when s more above than the substrate utilization constant (K_s), and this K_s value is numerically equal to substrate concentration when μ_g is half of μ_{max} .

Besides, the biodegradation kinetic also important in predicting the progress of contaminant biodegradation throughout the analysis of bioremediation processes. The biodegradation rate, k can be calculated using the first-order biodegradation kinetic model (equation 2) by assuming that population density and substrate concentration are the only factors which determine the degradation kinetics.

$$s = s_o \exp(-kt) \quad (2)$$

Moreover, the rate of contaminant removal or known as maximal contaminant specific degradation rate, Q_s can be measured by using the equation 3, which relating the mass of cells produced per mass of substrate utilized during the biodegradation process. Apparently, the higher the value of Q_s , the better the system in CB bioremediation is.

$$Q_s = \frac{\mu_g}{Y_{x/s}} \quad (3)$$

Whereas, s_0 is the initial substrate concentration, t is the time in hour, and $Y_{x/s}$ is the microbial yield in unit of gram cells produced per gram substrate utilized.

3. Results and Discussions

3.1. Microbial growth activity

CB utilization for growth by indigenous microbial consortia was studied intensively and measured spectrophotometrically as previously described. A comparison of the growth ability of these consortia on five different CB concentrations as sole source of carbon and energy in batch cultures for 6 days is shown in Figure 1.

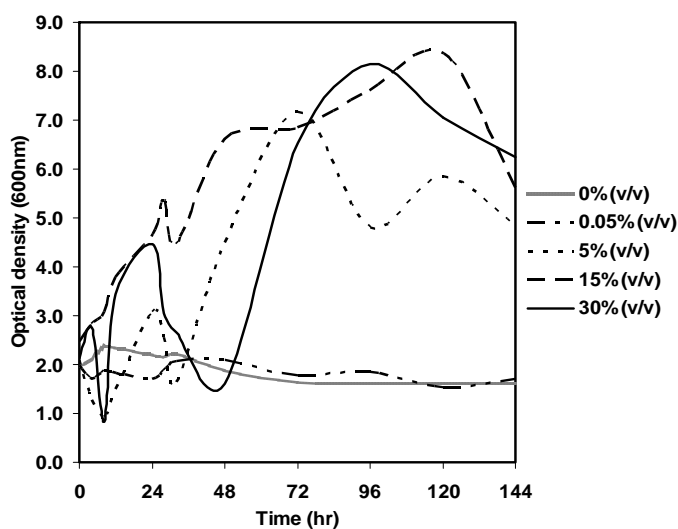


Figure 1. Growth of microbial consortia for 6 days at different initial CB concentrations

The initial CB levels were used at concentrations of 0.0% (v/v), 0.05% (v/v) or 0.0006mg/L, 5.0% (v/v) or 0.0553mg/L, 15.0% (v/v) or 0.1659mg/L and 30.0% (v/v) or 0.3317mg/L. These consortia were optimally grown under maintained aerobic conditions at room temperature and between pH 6 to 8.

The cells density for each CB concentrations increased due to increasing time as shown in Figure 1. The consortia of above 0.0553mg/L CB tested have rapidly increased and were using more than 50% of the substrate for growth activity. However, for both CB concentrations of 0.0 and 0.0553mg/L, the cells density

demonstrated the reverse patterns after 48 hours cultivation, then remained stable and slightly decreased until to 6 days.

In the shake flask, variations in biomass concentration were occurred in stages. As the CB concentrations released into the environment by industry vary enormously, the influence of an increased CB level on growth of indigenous consortia from local residential wastewater was tested. The results of microbial growth are summarized in Table 1.

Table 1. Influence of CB concentrations on growth of consortia in batch cultures; whereas X_0 is an initial biomass concentration and X_R is a residual biomass concentration

Initial CB levels (mg/L)	X_0 (mg/L), t=0	X_R (mg/L), after 3 days	X_R (mg/L), after 5 days	X_R (mg/L), after 6 days
0.0000	1956.00	1638.00	1620.00	1608.00
0.0006	1980.00	1788.00	1548.00	1698.00
0.0553	2094.00	7164.00	5850.00	4818.00
0.1659	2472.00	6846.00	8370.00	5610.00
0.3317	2112.00	6534.00	7044.00	6252.00

The results indicated an increase in the number of biomass concentration during batch cultivation for each CB concentrations tested excluding 0.0 and 0.0006mg/L CB. The specific growth rates, μ_g discovered for these consortia of 0.0553, 0.1659 and 0.3317mg/L CB tested were 0.0121, 0.0095 and 0.0079h⁻¹ respectively.

Nevertheless, the highest final biomass concentration after 6 days cultivation represents at the highest CB concentration tested. The population counts for these consortia have fast increasing within 5 days until the residual cells concentration reached 7044.0mg/L. Then, it started to decrease until the final biomass concentration of 6252.0mg/L. Therefore, the optimum growth activity for indigenous microbial consortia from local residential wastewater occurred around CB level of 0.3317mg/L during 6 days of experimentally implementation.

3.2. Efficiency of CB biodegradation

The successful of bioremediation process is often proved by the percent of reduction in contaminant other than measurement of increment of carbon dioxide released, increment of oxygen uptake and the presence of metabolic products. Furthermore, the ultimate goal for bioremediation itself shows the importance of mineralizing the contaminants, that is, to transform a hazardous chemical into harmless compounds such as carbon dioxide or other gas or inorganic substance, water, and cell material. Therefore, the reduction of CB toxic in cultures was measured as justification for the occurrence of CB biodegradation process in this study.

CB concentrations were introduced initially into shake flasks at different designed levels as prior discussed. Sterile condition and closed-system were ensured maintained during experimental works to avoid contamination and air stripping. Then, the reductions of CB concentrations were analyzed by HPLC analysis with a standard error of 15%. The results were extrapolated from the standard curve of the peak versus CB concentration to get the values in mg/L units.

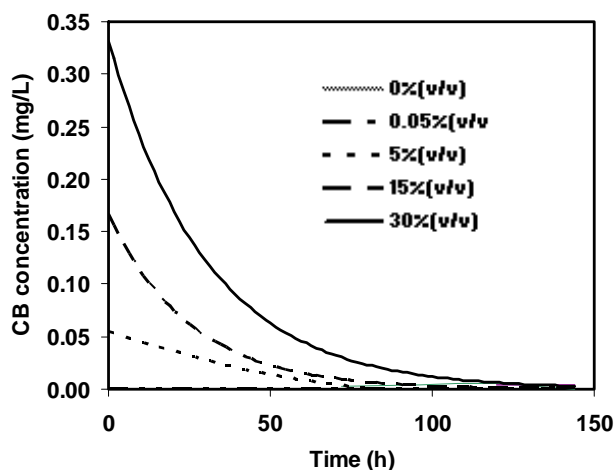


Figure 2. Efficiency of CB degradation for 6 days at different initial CB concentrations

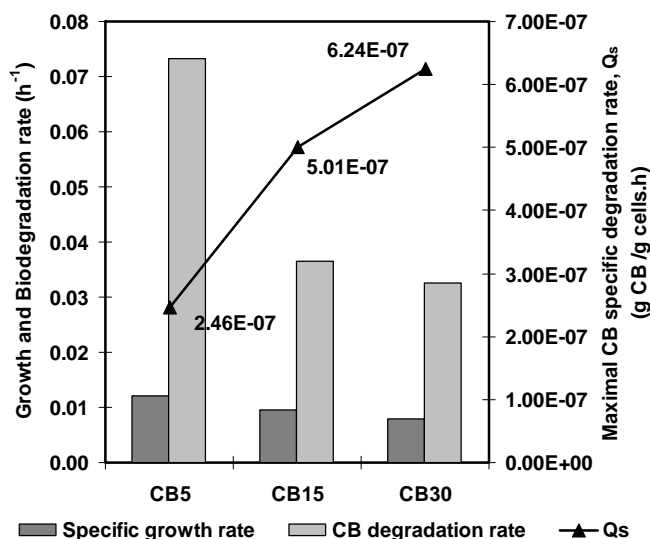


Figure 3. Summarized results of growth and degradation kinetics

Figure 2 illustrates a representative pattern of microbial CB biodegradation in flask cultures within 6 days. The CB concentrations occurred in the decrement trend for each initial CB concentration tested. Almost no CB was detected at the end of each culture. It is obviously shown that the consortia for CB concentration of 0.0553mg/L needed about 80 hours to approximately

complete the reduction of CB toxic in batch culture condition. Nevertheless, only 100 and about 144 hours were needed to degrade 0.1659 and 0.3317mg/L of CB respectively.

The consortia were tested for its ability to degrade different concentrations of CB by evaluating the growth and biodegradation kinetic parameters as summarized in Figure 3. Therefore, the highest efficiency of CB degradation was crucial in this study. This efficiency of CB biodegradation was referred to Q_s , which has been previously described.

From Figure 3, it is obviously shown that the highest Q_s value was at the highest initial CB concentration tested even though the highest specific growth and biodegradation rate was at 0.0553mg/L CB. This Q_s value for CB concentration of 0.3317mg/L is 6.24×10^{-7} gram CB utilized per gram cells produced per hour. Consequently, the optimal efficiency of CB degradation in this study occurred at the highest CB concentration used, namely 0.3317mg/L.

The disappearance of CB from culture medium had a specific, almost disproportional-like profile when transformed by this diversity of indigenous consortia. This phenomenon was occurred for each culture and dropped outstandingly at the end of the cultivation, i.e. after 6 days. The results discovered in this study agreed well with other researchers, which the initial CB concentrations tested were in the ranges of validated levels from other researchers as represents in Table 3.

Apparently, in the field application, CB removal is dependent on many variables other than substrate factor, specifically the effect of initial concentration. They includes the environmental and microbiological factors such as [3]; predation by other microorganisms, bioavailability of the contaminant, presence of preferential carbon sources, and need for an acclimatization period.

4. Conclusions

The results revealed in this study agreed well with the well-known fact that initial levels of toxic compound at below threshold value were prominent for biodegradation. The CB degradation significantly depends upon the initial CB concentration although the microbial consortia utilized did not have previously been exposed to CB. It is evidently emphasized that low toxic concentrations did not provide sufficient energy for growth.

In addition, the indigenous microbial consortia originated from local residential wastewater are able to degrade CB if provided with appropriate conditions. On the other hand, the consortia were robust enough to grow on recalcitrant toxic aromatic compound without any prior adaptation. The outcomes of this study provide useful information and guideline in employing the consortia from residential wastewater to solve the chlorobenzene contaminated in the environment. Further studies of CB biodegradation in bioreactor are still in progress. This study focused more on the performance of consortia under continuous culture conditions

Table 3. Comparison of the results on effect of CB levels with other researchers

Author	Microbes	Range of CB levels (mg/L)	Remark
Reineke and Knackmuss [26]	CB-degrading bacterium	35.71 to 371.43	-max growth rate at 35.71mg/L and above -above 35.71mg/L, growth increase but disturbed exponential
Yadav <i>et al.</i> [27]	<i>Phanerochaete chrysosporium</i>	0.50 to 100.00	-observation within 7 days -optimum degradation at 5.0mg/L with 95.4 ± 0.2 % CB degrade and final cell dry weight at 86.7 ± 4.4 mg
Zaitsev <i>et al.</i> [28]	<i>Rhodococcus opacus</i> GM-14	100.00 to 600.00	-degradation for 7 days -between 300 to 500mg/L, prolonged the lag period but growth increase -no growth at 600mg/L
Sommer and Gorisch [29]	<i>Xanthobacter flavus</i> 14pl	0.00 to 250.00	-growth increase with increasing CB -started inhibitory at above 142.86mg/L
Kiernicka <i>et al.</i> [30]	<i>Escherichia hermanii</i>	202.00 to 394.00	-nearly complete after 350 hours -degradation occurred stepwise -highest CB has negative effect on growth
This study	Microbial consortia	0.00 to 0.33	-no growth below 0.0006mg/L -specific degradation rate increase due to increasing initial CB concentration

Acknowledgements

The financial support of the UTM-PTP, SPS, UTM Skudai for scholarship of postgraduate students for one of the author is gratefully acknowledged. The author wish to thank Department of Bioprocess Engineering, UTM Skudai. The technicians in Bioprocess laboratory are also greatly appreciated. This research was financially supported by UTM Short Term Grant (vote: 75200).

References

- [1] Cookson, John T., Jr., 1995. Bioremediation Engineering; Design and Application. McGraw-Hill, Inc.
- [2] Baker, K. H.; and Diane S. H., 1994. Bioremediation. McGraw-Hill, Inc.
- [3] Eweis, J. B.; Ergas, S. J.; Chang, D. P. Y.; and Schroeder, E. D., 1998. Bioremediation Principles, McGraw-Hill, Inc.
- [4] U. S. Environmental Protection Agency, 1980. Ambient water quality criteria for chlorinated benzenes. EPA 400/5-80-028. Office of Water Regulation and Standards, Washington, D. C.
- [5] Harris, J.; Coons, S.; Byrne, M.; Fiksel, J.; and Korte, F., 1985. An exposure and risk assessment for dichlorobenzenes. EPA 440/4-81-019. Office of Water Regulations and Standards, U. S. Environmental Protection Agency, Washington, D. C.
- [6] Ding, W.H., Aldous, K.M., Briggs, R.G., 1992. Application of multivariate statistical analysis to evaluate local sources of chlorobenzene congeners in soil samples. *Chemosphere*, 25:675-690.
- [7] Boyd, E.M., Killham, K., Wright, J., 1997. Toxicity assesment of xenobiotic contaminated groundwater using *lux* modified *Pseudomonas fluorescens*. *Chemospre*, 35:1967-1985.
- [8] Rogers, H.R., Campbell, J.A., Crathorne, B., Dobbs, A.J., 1989a. The occurrence of chlorobenzenes and permethrins in twelve UK sewage sludges. *Water Res.*, 23:913-921.
- [9] Wang, M.J., McGrath, S.P., Jones, K.C., 1992. The chlorobenzene content of archived sewage sludges. *Sci. Total Environment.*, 121:159-175.
- [10] Masunaga, S., Yonezawa, Y., Urushigawa, Y., 1991. The distribution of chlorobenzenes in the bottom sediments of Ise Bay, *Water Res.*, 25:275-288.
- [11] Lee, C. L., Fange, M.D., 1997. Sources and distribution of chlorobenzenes and hexachlorobutadiene in surficial sediments along coast

- of south-western Taiwan. *Chemosphere*. 35:2039-2050.
- [12] Rogers, H.R., Crathorne, B., Leatherland, T.M., 1989b. Occurrence of chlorobenzene isomers in the water column of a UK estuary. *Mar. Pollut. Bull.*, 20:276-281.
- [13] Harper, D.J., Ridgeway, I.M., Leatherland, T.M., 1992. Concentrations of hexachlorobenzene, trichlorobenzenes and chloroform in the waters of the Forth estuary, Scotland. *Mar. Pollut. Bull.*, 24:244-249.
- [14] Meharg, A.A., Wright, J., Leeks, G.J.L., Wass, P.D., Osborn, D., 2000. Spatial and temporal regulation of the pesticide dieldrin within industrial catchments. *Sci. Total Environ.* 251/252:255-263.
- [15] Bittkau, A.; Geyer, R.; Bhatt, M.; and Schlosser, D., 2004. Enhancement of biodegradability of aromatic groundwater contaminants. *Toxicology*, 205:201-210.
- [16] Schraa G, Boone ML, Jetten MSM, Van Neerven ARW, Colberg PJ, Zehnder AJB, 1986. Degradation of 1,4-dichlorobenzene by *Alcaligenes* sp. Strain A175. *Applied Environmental Microbiology*, 52(6):1374-1381
- [17] Spain, J. C.; and Nishino, S. F., 1987. Degradation of 1,4-dichlorobenzene by a *Pseudomonas* sp. *Applied and Environmental Microbiology*, 53:3-19.
- [18] Haigler, B. E.; Pettigrew, C. A.; and Spain, J. C., 1992. Biodegradation of mixtures of benzenes by *Pseudomonas* sp. Strain JS150. *Applied and Environmental Microbiology* 58:2237-2244.
- [19] Aelion, C. M.; Swindoll, C. M.; and Pfaender, F. K., 1987. Adaptation to and Biodegradation of xenobiotic compounds by microbial communities from a Pristine Aquifer. *Applied and Environmental Microbiology* 59(9):2212-2217.
- [20] Nishino SF, Spain JC, Pettigrew CA, 1994. Biodegradation of chlorobenzene by indigenous bacteria. *Environmental Toxicology and Chemistry*, 13(6):871-877.
- [21] Kao, C. M.; and Presser, J., 1999. Intrinsic bioremediation of trichloroethylene and chlorobenzene: field and laboratory studies. *Journal of Hazardous Materials*, B(69):67-79.
- [22] Balcke, G. U.; Turunen, L. P.; Geyer, R.; Wenderoth, D. F.; and Schlosser, D., 2004. Chlorobenzene biodegradation under consecutive aerobic-anaerobic conditions. *FEMS Microbiology Ecology*, 49(1):109-120.
- [23] Dilmeghani, M.; and Zahir, K. O., 2000. Kinetics and mechanism of chlorobenzene degradation in aqueous samples using advanced oxidation process. *Journal of Environmental Quality*, 30:2062-2070.
- [24] Monod, J.; 1942. Recherches sur la croissance des cultures bacteriennes, Hermann et Cie., Paris.
- [25] Pirt, S., J., 1975. Principles of Microbe and Cell Cultivation. Blackwell, London.
- [26] Reineke W, Knackmuss HJ, 1984. Microbial metabolism of haloaromatics: isolation and properties of a chlorobenzene-degrading bacterium. *Applied Environmental Microbiology*, 47(2):395-402.
- [27] Yadav, J. S.; Wallace, R. E.; and Reddt, A., 1995. Mineralization of mono- and dichlorobenzenes and simultaneous by the White rot Fungus *Phanerochaete chrysosporium*. *Applied and Environmental Microbiology* 61(2):677-680.
- [28] Zaitsev, G. M.; Uotila, J. S.; Tsitko, I. V.; Lobanok, A. G.; and Salkinoja-Salonen, M. S., 1995. Utilization of Halogenated Benzenes, Phenols and benzoates by *Rhodococcus opacus* GM-14. *Applied and Environmental Microbiology* 61(12):4191-4201.
- [29] Sommer, C.; and Gorisch, H., 1997. Enzymology of the degradation of (di)chlorobenzene by *Xanthobacter flavus* 14pl. *Arch Microbiol*, 167:384-391.
- [30] Kiernicka J, Seignez C, Peringer P, 1999. *Escherichia hermannii* – a new bacterial strain for chlorobenzene degradation. *Letters in Applied Microbiology*, 28(1):27-30.

Co-Generation of Synthesis Gas and Higher Hydrocarbon Over Catalytic Dielectric Barrier Discharge Plasma Reactor

Istadi¹, and Nor Aishah Saidina Amin²

¹ Permanent Address: Chemical Reaction Engineering and Catalysis (CREC) Group,
Department of Chemical Engineering, Diponegoro University, Jln. Prof. Sudharto, Semarang, Indonesia 50239
E-mail: istadi@tekim.ft.undip.ac.id

² Chemical Reaction Engineering Group (CREG), Faculty of Chemical and Natural Resources Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia

Abstract

Potential application of hybrid catalytic-DBD plasma reactor for co-generation of C₂₊ hydrocarbons and synthesis gases from methane and carbon dioxide is highlighted. The synergism of the catalyst, feed ratio and the plasma discharge affect predominantly the products distribution, particularly C₂₊ hydrocarbons selectivity. The CH₄/CO₂ feed ratio, total feed flow rate, and discharge voltage exhibit significant effects on the reactor performances. However, increasing the reactor wall temperature has no apparent influence on the selectivity to C₂₊ hydrocarbons and hydrogen within the investigated range. The hybrid catalytic-DBD plasma reactor is potential for co-generation of higher hydrocarbons and synthesis gas rather than conventional catalytic reactor over CaO-MnO/CeO₂ catalyst.

Keywords:

plasma reactor; dielectric-barrier discharge; synthesis gas; higher hydrocarbons; methane conversion

1. Introduction

Among all greenhouse gases, CH₄ and CO₂ contribute most to the man-made greenhouse effect. Any success in the research and development of a feasible utilization of CH₄ and CO₂ will signify the attainment of double objectives of slowing down a buildup of greenhouse gases in the atmosphere and better carbon resource utilization. A possible utilization of CH₄ and CO₂ is the co-generation of higher hydrocarbons (HHCs) and synthesis gas. However, there exist considerable difficulties in such utilization of CH₄ and CO₂ in an economical way. In this paper, the combined action of catalysts and a non-equilibrium gas discharge may lead to alternative methods of syngas and hydrocarbon production from CH₄ and CO₂.

Non-thermal plasma can be defined as a gas consisting of electrons, highly excited atoms and molecules, ions, radicals, photons and neutral particles in which the electrons have much higher energy than the neutral gas particles. The energetic electrons collide with molecules in the gas, resulting in excitation, ionization, electron multiplication, and formation of atoms and metastable compounds (Caldwell et al., 2001; Larkin et al., 2001; Kogelschatz, 2003). When the electric field in the discharge gap is high enough to cause breakdown in most gases a large number of microdischarges are observed. The active atoms and metastable compounds subsequently collide with molecules and reactions may occur. Due to high energy

electron of the plasma reactor, it is expected that methane and carbon dioxide gases could be activated easily in the plasma environment and is converted into synthesis gases and higher hydrocarbons. It is expected that non-conventional Dielectric Barrier Discharge (DBD) plasma reactor is an efficient tool for converting the greenhouse gases, CH₄ and CO₂, to synthesis gas and higher hydrocarbons at low temperature and ambient pressure (Caldwell et al., 2001; Larkin et al., 2001; Liu et al., 1999; Zou et al., 2003).

The present contribution is intended to study the potential application of DBD plasma reactor for methane and carbon dioxide conversion. The effect of operating parameters on the plasma reactor performance is also investigated based on the simulation of artificial neural network – based model.

2. Experimentals

2.1. Configuration of dielectric barrier discharge plasma reactor

The schematic diagram of experimental rig of DBD plasma reactor is depicted in Figure 1. Figure 2 presents the detail schematic diagram of DBD plasma reactor design. A high voltage pulse AC generator supplies a voltage from 0 kV to 17.5 kV with a pulsed waveform at a frequency of up to 10 kHz was used. The voltage measurement was

conducted using an oscilloscope (ISO-TECH ISR 622) equipped with a high voltage probe (manufactured by Atama Tech Sdn. Bhd.) which was calibrated by Tektronix P6015 probe. Methane and carbon dioxide were fed to the reactor at a total flow rate based on design of experiment. The amount of catalyst (5 g) was put in the discharge zone of the DBD plasma reactor. Pertaining to catalytic DBD plasma reactor, the reactor wall temperature was maintained by a tube furnace (CARBOLITE), while the gas flow rates were measured and controlled by volumetric flow controllers (Alicat Scientific, Inc.). The products and unreacted gases were analyzed by an online gas chromatography (HP AGILENT 6890 series) equipped with a thermal conductivity detector (TCD), Molecular Sieve 5A and Porapak N packed columns.

2.2. Catalyst preparation

The catalyst (12.8 wt.% CaO-6.4 wt.% MnO/CeO₂) used in this catalytic DBD plasma reactor is based on the optimization results over conventional fixed bed reactor as described elsewhere (Istadi and Amin, 2005, 2006a). The ternary CaO-MnO/CeO₂ metal oxide catalyst was prepared by simultaneous incipient wetness impregnation method. Powdered CeO₂ (MERCK) was first immersed into an aqueous mixed solutions of Ca(NO₃)₂·4H₂O (99%, ALDRICH) and Mn(NO₃)₂·4H₂O (98.5%, FLUKA) for 6 h at ambient temperature and was dried overnight in the oven (MEMMERT) at 393 K. Next, the powder was calcined at 1123 K in a muffle furnace (CARBOLITE) for 10 h and then crushed into the desired size (45-60 mesh).

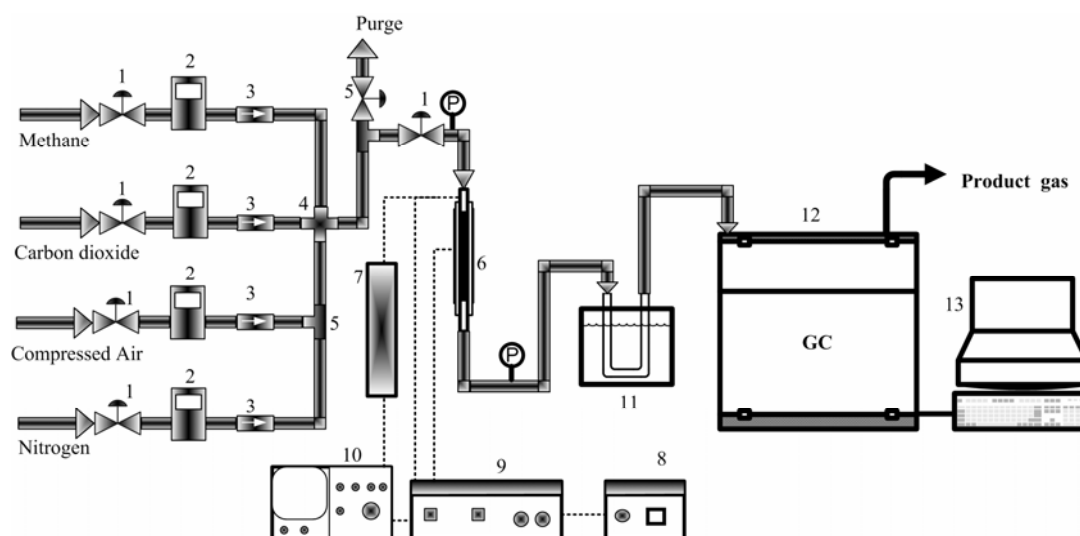
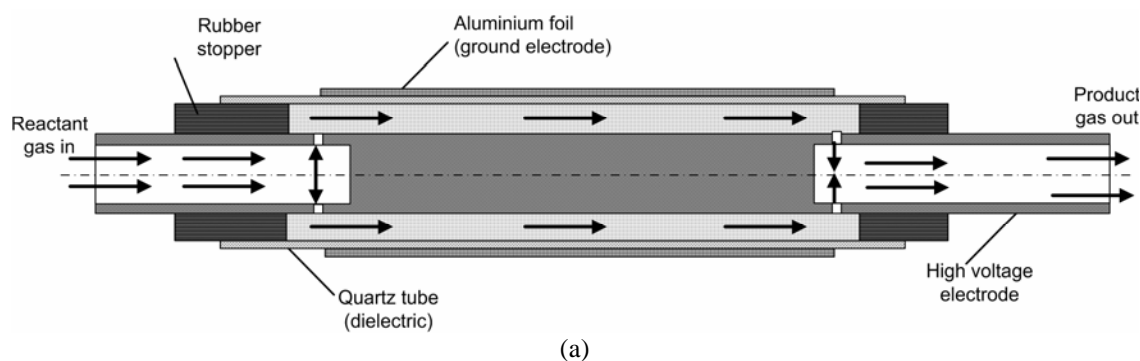
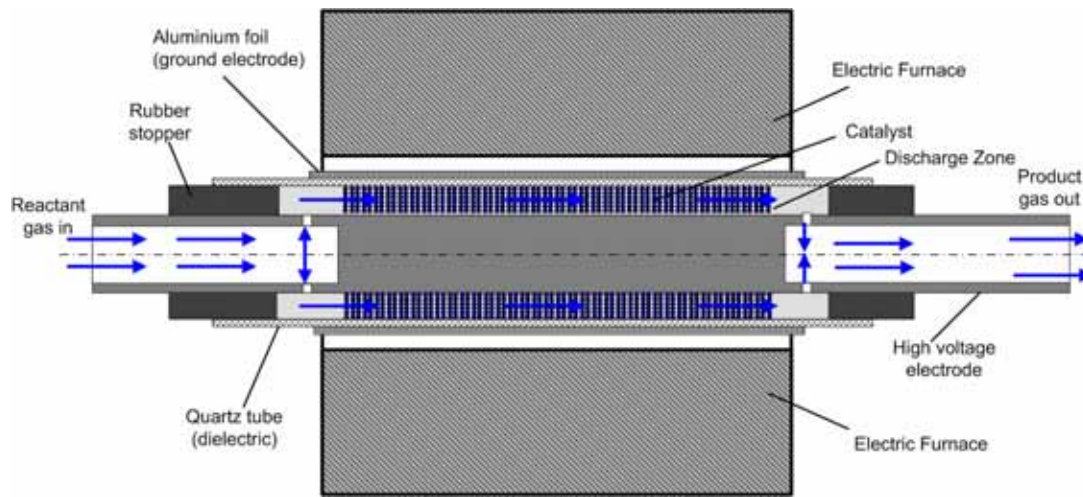


Figure 1. Schematic diagram for experimental rig set-up of DBD plasma reactor: (1). Ball valve; (2). Volumetric flow controller; (3). Check valve; (4,5). Four and three way valves; (6). DBD plasma reactor; (7). High voltage probe; (8). DC power supply; (9). High voltage AC generator; (10). Oscilloscope; (11). Condensor; (12). Online GC; (13). Computer for GC; (P). Pressure gage





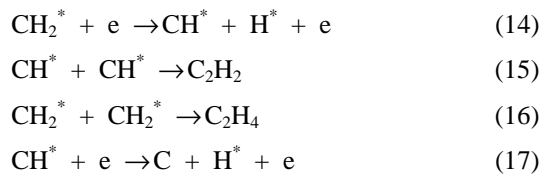
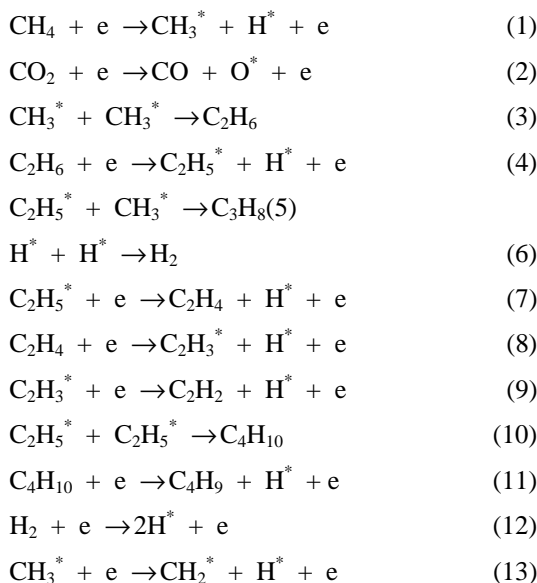
(b)

Figure 2. Schematic diagram of DBD plasma reactor: (a) without catalyst and heating, (b) with catalyst and heating

3. Results and Discussion

3.1. Possible Reaction Mechanism of CH₄ and CO₂ Conversion in Hybrid Catalytic – DBD Plasma Reactor

In plasma reactor, the energetic electrons collide with molecules in the gas phase resulting in excitation, ionization, electron multiplication, and the formation of atoms and metastable compounds. When the electric field in the discharge gap is high enough to cause breakdowns in most gases, a large number of microdischarges are observed. The active atoms and metastable compounds subsequently collide with molecules, and reactions may occur. For CO₂ reforming of methane, it is expected that CH₄ and CO₂ participate in the reactions and are converted to higher hydrocarbons and synthesis gases. The possible reactions occur in the CH₄ and CO₂ conversions over the DBD plasma reactor without and with catalyst at low temperature can be described as follows:



In the reactions, e denotes high energy electron, while * expresses radical species from dissociation reaction.

3.2. Effect of operating parameters on performances of DBD plasma reactor

This section presents the effect of operating parameters in catalytic DBD plasma reactor for CH₄ and CO₂ conversion resulted from the ANN simulation. The model validation was performed by training the network model with experimental data designed by Central Composite Design (CCD) (Istadi 2006). The ANN-based model has a good fitting toward the experimental data. The simulations were carried out by varying one operating parameter, while the remaining parameters are kept constant. In fact, CH₄ conversion, C₂₊ hydrocarbons selectivity and yield, H₂ selectivity, and H₂/CO ratio are affected by CH₄/CO₂ feed ratio, discharge voltage, total feed flow rate and reactor wall temperature as depicted in Figures 3 - 6 developed from the ANN-based model simulation.

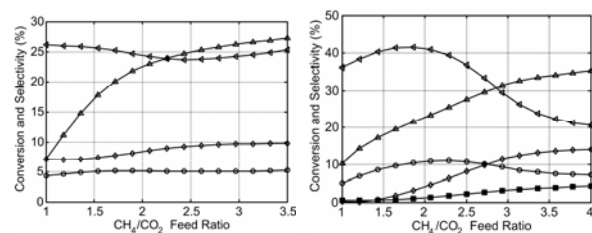


Figure 3. Effect of CH₄/CO₂ feed ratio on catalytic DBD plasma reactor performance at discharge voltage 15 kV, 30 cm³/min total feed flow rate and reactor temperature 200 °C. ◀ CH₄ conversion, ▲ C₂₊ selectivity, ◇ H₂ selectivity, ⊖ C₂₊ yield, ■ H₂/CO ratio

Figure 3 presents the simulation of CH₄/CO₂ feed ratio effect on the catalytic DBD plasma process performance. Increasing the CH₄ concentration in the feed favours the selectivity of C₂₊ hydrocarbons and hydrogen significantly, but the C₂₊ hydrocarbons yield is only slightly affected due to the decrease of CH₄ conversion. It is suggested that the CH₄ concentration in the feed is an important factor for the total amount of hydrocarbons produced. However, increasing CH₄/CO₂ ratio to 4 reduces the methane conversion considerably and leads to enhanced C₂₊ hydrocarbons selectivity and H₂/CO ratio. Increasing CO₂ concentration in the feed improves the CH₄ conversion and CO selectivity which may be due to promoting the CH₄ conversion by oxygen from CO₂ decomposition to oxygen active species. This phenomenon is corroborated with the results of Zhang et al. (2001). The yield of gaseous hydrocarbons (C₂₊) increases slightly with the CH₄/CO₂ feed ratio as exhibited in Figure 3 and consequently lowers at higher ratio. It is possible to control the composition of C₂₊ hydrocarbons and hydrogen products by adjusting the CH₄/CO₂ feed ratio. From the ANN simulation result in Figure 3, it can be shown that increasing CH₄/CO₂ ratio above 2.5 exhibits low enhancement of C₂₊ yield and lower CH₄ conversion. In this work, composition of the feed gas (CH₄/CO₂ ratio) is an essential factor to adjust the product distribution. Obviously, more methane in the feed will produce more light hydrocarbons.

Varying the discharge power/voltage affects predominantly on methane conversion and higher hydrocarbons (C₂-C₃) selectivity. The methane conversion increases with discharge voltage as exhibited in Figure 4. More plasma species may be generated at higher discharge voltage. Previous researchers suggested that the conversions of CH₄ and CO₂ were enhanced with discharge power in a catalytic DBD plasma reactor (Caldwell et al., 2001; Eliasson et al., 2000; Zhang et al., 2001, 2002). From Figure 4, the selectivities of C₂₊ hydrocarbons and hydrogen decrease slightly with the discharge voltage which is corroborated with the results of Liu et al. (2001). This means that increasing discharge power may destroy the light hydrocarbons (C₂-C₃). In comparison, high C₂ and C₃ hydrocarbons selectivities of 36.4% and 18% respectively were reported by Eliasson et al. (2000) at discharge power 200 W (30 kHz) using DBD plasma reactor with zeolite catalyst. In this research, the lower range of discharge power (discharge voltage 12 - 17 kV and frequency 2 kHz) does not improve the H₂ selectivity over DBD plasma reactor although the catalyst and the heating was introduced in the discharge space. Higher discharge power is suggested to be efficient for methane conversion. As the discharge power increases, the bulk gas temperature in the reaction zone may also increase.

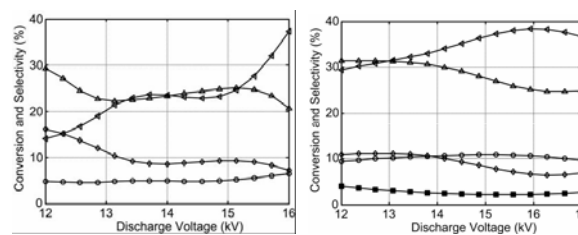


Figure 4. Effect of discharge voltage on catalytic DBD plasma reactor performance at CH₄/CO₂ feed ratio 2.5, 30 cm³/min total feed flow rate and reactor temperature 200 °C. ◀ CH₄ conversion, ▲ C₂₊ selectivity, ◇ H₂ selectivity, ⊖ C₂₊ yield, ■ H₂/CO ratio

The total feed flow rate also affects predominantly on the residence time of gases within the discharge zone in the catalytic DBD plasma reactor which consequently influences the collisions between the gas molecules and the energetic electrons. Increasing the total feed flow rate reduces the residence time of gases and therefore decreases the methane conversion quickly as demonstrated in Figure 5. A lower feed flow rate is beneficial for producing high yields light hydrocarbons (C₂₊) and synthesis gases with higher H₂/CO ratio as reported by Li et al. (2004). From Figure 5, it is shown that increasing the total feed flow rate decreases the CH₄ conversion markedly and affects the C₂₊ hydrocarbons selectivity slightly. The hydrogen selectivity is also small affected by the total feed flow rate within the range of operating conditions. Indeed, the total feed flow rate affects significantly on the methane conversion rather than selectivity of C₂₊ hydrocarbons and hydrogen. Actually, the low total feed flow rate (high residence time) leads to high intimate collision among the gas molecules, the catalyst and high energetic electrons. The high intensive collisions favour the methane and carbon dioxide conversions to C₂₊ hydrocarbons.

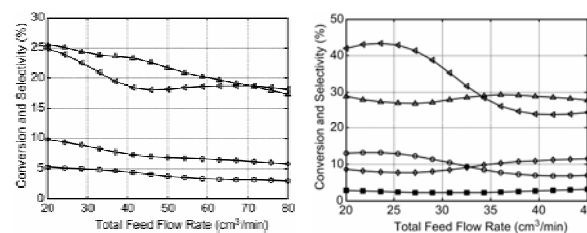


Figure 5. Effect of total feed flow rate on catalytic DBD plasma reactor performance at CH₄/CO₂ feed ratio 2.5, discharge voltage 15 kV and reactor temperature 200 °C. ◀ CH₄ conversion, ▲ C₂₊ selectivity, ◇ H₂ selectivity, ⊖ C₂₊ yield, ■ H₂/CO ratio

Pertaining to the reactor wall temperature, Figure 6 presents the effect of reactor temperature variation on the performance of catalytic DBD plasma reactor. Thermodynamic equilibrium calculations demonstrated that normal chemical reactions between CH₄ and CO₂ cannot be expected at temperatures lower than 523 K (Istadi and Amin, 2005b). In endothermic reactions, normally high temperatures are required to add enthalpy. In this research, the methane and carbon dioxide reaction over CaO-

MnO/CeO₂ catalyst in the DBD plasma reactor is influenced significantly by energetic electrons. The carbon was also formed at the entire surface of electrode during the reaction.

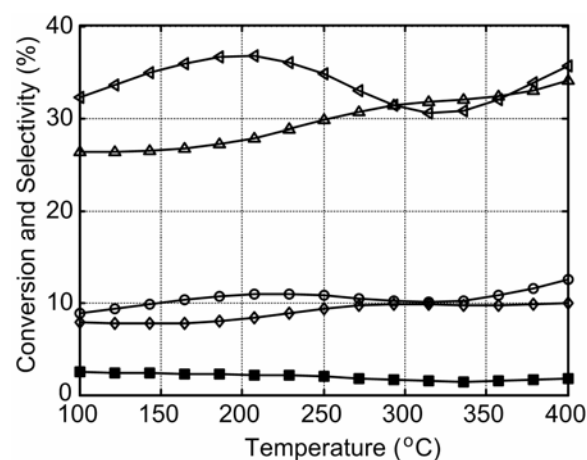


Figure 6. Effect of reactor wall temperature on catalytic DBD plasma reactor performance at CH₄/CO₂ feed ratio 2.5, discharge voltage 15 kV and total feed flow rate 30 cm³/min. \triangleleft CH₄ conversion, \triangle C₂₊ selectivity, \diamond H₂ selectivity, \circ C₂₊ yield, \blacksquare H₂/CO ratio

3.3. Effect of heterogeneous catalysis on performances of DBD plasma reactor

The catalyst located in the discharge gap can increase the time and area of contact in addition to other modification of electronic properties. Through the hybrid system, the chemisorption and desorption performances of the catalyst may be modified in the catalyst surface which is dependent on the amount and concentration of surface charge and the species on the catalyst surface (Kim et al., 2004). The results enhancement was also reported by Eliasson et al. (2000) over DBD plasma reactor with high input power of 500 W (20 kV and 30 kHz) that the zeolite catalyst introduction significantly increased the selectivity of light hydrocarbons compared to that in the absence of the zeolite, i.e. C₂H₆ selectivity from 8.5 to 11.1% and decreased selectivities of C₅₊ and other oxygenates from 41.2 to 34.2%. The CO selectivity slightly decreased when using the zeolite, but the formations of carbon and plasma polymerization were inhibited.

A good plasma catalytic activity is achieved only when streamer discharges are present. At higher temperatures, the streamer discharge may turn to an arc-like discharge and thermal effects dominate the reactions. The non-equilibrium streamer discharge at lower temperatures favoured the formation of the higher hydrocarbons. From Figure 7, it is evident that the current range of reactor temperature (100-400 °C) only affects the catalytic DBD plasma reactor slightly. The methane conversion is very slightly affected by reactor wall temperature over the CaO-MnO/CeO₂ catalyst. The C₂₊ hydrocarbons selectivity is enhanced slightly by the reactor temperature which may be due to the altering the catalyst surface phenomena. The

adsorption-desorption, heterogeneous catalytic and electronic properties of the catalysts may change the surface reaction activity when electrically charged. However, the chemistry and physical phenomena at the catalyst surface can not be determined in the sense of traditional catalyst.

Indeed, the roles of catalyst in the plasma process are: (a) increasing the time and area of contact (reaction surface area), (b) maintaining and probably increasing the non-equilibrium properties of gas discharge, (c) acting as a dielectric material, and (d) improving the selectivity and efficiency of plasma processes by surface reactions.

4. Conclusions

It is found that a hybrid catalytic DBD plasma reactor can be potentially used for the co-generation of C₂₊ hydrocarbons (ethane, ethylene, acetylene, and propane) and synthesis gases from methane and carbon dioxide. The synergism of the catalyst (CaO-MnO/CeO₂), feed ratio and the plasma discharge affect the products distribution, particularly C₂₊ hydrocarbons selectivity. The introduction of catalyst in the discharge space improved the performance the CH₄ and CO₂ conversion since the time and area of contact are enhanced in addition to other modification of the electronic properties. It can be concluded that the hybrid catalytic DBD plasma reactor is more suitable for CO₂ OCM process than the conventional catalytic reactor over CaO-MnO/CeO₂ catalyst.

Acknowledgement

The authors would like to express their sincere gratitude to the Ministry of Science, Technology and Innovation, (MOSTI) Malaysia for the financial support received under the Project No 02-02-06-0016 EA099

References

- [1]. Caldwell, T.A., Le, H., Lobban, L.L., Mallinson, R.G. 2001. Partial Oxidation of Methane to Form Synthesis Gas in a Tubular AC Plasma Reactor, in Spivey, J.J.; Iglesia, E.; Fleisch, T.H. Eds. *Stud. Surf. Sci. Catal.*, 136, Elsevier B.V., Amsterdam.
- [2]. Eliasson, B., Liu, C.J., Kogelschatz, U. 2000. Direct Conversion of Methane and Carbon Dioxide to Higher Hydrocarbons using Catalytic Dielectric-Barrier Discharges with Zeolites. *Ind. Eng. Chem. Res.* 39: 1221-1227.
- [3]. Istadi, and Amin, N.A.S. 2005a. A Hybrid Numerical Approach for Multi-Responses Optimization of Process Parameters and Catalyst Compositions in CO₂ OCM Process over CaO-MnO/CeO₂ Catalyst. *Chem. Eng. J.* 106: 213-227.
- [4]. Istadi, and Amin, N.A.S. 2005b. A Thermodynamic Analysis of Co-generation of C₂ Hydrocarbons and Synthesis Gases from Methane and Carbon Dioxide by Direct Gibbs Free Energy Minimization. *J. Nat. Gas Chem.* 14: 140-150.

- [5]. Istadi, and Amin, N.A.S. 2006a. Optimization of Process Parameters and Catalyst Compositions in Carbon Dioxide Oxidative Coupling of Methane over CaO-MnO/CeO₂ Catalyst using Response Surface Methodology. *Fuel Process. Technol.*, 87: 449-459.
- [6]. Istadi. 2006. Catalytic Conversion of Methane and Carbon Dioxide in A Conventional Fixed Bed and Dielectric-Barrier Discharge Plasma Reactors. *PhD Thesis*. Universiti Teknologi Malaysia, Johor, Malaysia.
- [7]. Kim, S.S., Lee, H., Na, B.K., and Song, H.K. 2004. Plasma-assisted Reduction of Supported Metal Catalyst using Atmospheric Dielectric-barrier Discharge. *Catal. Today* 89: 193-200.
- [8]. Kogelschatz, U. 2003. Dielectric-barrier Discharges: Their History, Discharge Physics, and Industrial Applications. *Plasma Chem. Plasma Process.* 23: 1-46.
- [9]. Larkin, D.W., Zhou, L., Lobban, L.L., Mallinson, R.G. 2001. Product Selectivity Control and Organic Oxygenate Pathways from Partial Oxidation of Methane in a Silent Electric Discharge Reactor. *Ind. Eng. Chem. Res.* 40: 5496-5506.
- [10]. Li, M.W., Xu, G.H., Tian, Y.L., Chen, L., and Fu, H.F. 2004. Carbon Dioxide Reforming of Methane Using DC Corona Discharge Plasma Reaction. *J. Phys. Chem. A* 108: 1687-1693.
- [11]. Liu, C.J., Xue, B., Eliasson, B., He, F., Li, Y. and Xu, G.H. 2001. Methane Conversion to Higher Hydrocarbons in the Presence of Carbon Dioxide using Dielectric Barrier-Discharge Plasmas. *Plasma Chem. Plasma Process.* 21: 301-309.
- [12]. Liu, C.-J., Xu, G.-H., Wang, T. 1999. Non-Thermal Plasma Approaches in CO₂ Utilization. *Fuel Process. Technol.* 58: 119-134.
- [13]. Zhang, K., Eliasson, B., and Kogelschatz, U. 2002. Direct Conversion of Greenhouse Gases to Synthesis Gas and C₄ Hydrocarbons over Zeolite HY Promoted by a Dielectric-Barrier Discharge. *Ind. Eng. Chem. Res.* 41: 1462-1468.
- [14]. Zhang, K., Kogelschatz, U. and Eliasson, B. 2001. Conversion of Greenhouse Gases to Synthesis Gas and Higher Hydrocarbons. *Energy Fuels* 15: 395-402.
- [15]. Zou, J.-J., Zhang, Y., Liu, C.-J., Li, Y., Eliasson, B. 2003. Starch-enhanced Synthesis of Oxygenates from Methane and Carbon Dioxide Using Dielectric-barrier Discharges. *Plasma Chem. Plasma Process.* 23: 69-82.

Production and Characterization of crude chitinase from *Trichoderma virens*

Azaliza Safarida Wasli¹, Madihah Md Salleh^{1*}, Rosli Md Illias²

¹Biology Department, Faculty of Science

University Technology Malaysia, 81310, Skudai, Johor Darul Takzim

Tel: +60-7-5532526, Email: azasafarida@yahoo.com

²Bioprocess Department, Faculty of Chemical Engineering and Natural Resources

University Technology Malaysia, 81310, Skudai, Johor Darul Takzim

Abstract

Chitinase hydrolyzes chitin, a biopolymer of *N*-acetylglucosamine which is being widely used in biological and agricultural research. Optimization of chitinase production by *Trichoderma virens* was conducted using batch cultures of *Trichoderma Minimal Medium* (TM medium) that consists of 2.0% (w/v) colloidal chitin as substrate. Fermentation was performed at agitation speed of 200 rpm and incubated at 30°C in 250 ml shake flasks. Optimum chitin concentration was obtained at 4.0% (w/v). The addition of a stimulant, methanol, at 0.2% (v/v) enhances chitinase production by 71.4%. Higher concentrations of methanol would inhibit chitinase production. Works for optimizing agitation speed was carried out in a 2.0 L batch culture using *B. Braun Biostat-B* bioreactor. The optimum conditions for chitinase productions were performed in the same bioreactor using disc turbine impellers (diameter of 5.5 cm) with air flow rate of 1.0 l/m and agitation speed of 250 rpm resulting in an increased chitinase activity of 0.447 U/ml compared to 0.052 U/ml in shake flask fermentation. This improvement of 88.4% might be due to efficient mixing and high oxygen transfer rate which is essential for fungal growth. Characterization of crude chitinase was carried out to determine the optimal conditions for the crude chitinase activity to hydrolyze chitin and the crude enzyme stability. The crucial physical factors affecting the optimum and stability of chitinase activity were pH and temperature. pH optimum was achieved at pH 4.0 in citrate phosphate buffer and chitinase was very stable at pH 3.0-7.0, with 350% of residual activity is retained at pH 7.0. Optimum temperature for chitinase activity was achieved at 50°C while its stability decreases with the increment of temperature. Only 37.5% of residual chitinase activity was retained at 70°C.

Keywords : Chitinase, *Trichoderma virens*, batch culture, optimization, characterization

1. Introduction

Chitinases (E.C. 3.2.1.14) are a group of enzymes that is responsible of hydrolyzing chitin polymer into its oligomers or monomers [1]. Chitinases are naturally produced by bacteria, yeasts, protozoa, plants, humans and animals where the enzyme is usually involved in natural protective mechanism [1-4]. Chitin polymer is the second most abundant naturally occurring polymer after cellulose and it exist as a structural polysaccharide of β -1,4-*N*-acetyl-D- glucosamine residues. This polymer is mainly derived from marine invertebrates such as shrimp, crabs, oysters and cuttlefishes. Chitin and its derivatives produced from enzyme hydrolysis are being widely

applied in the field of medicine, agriculture, food industry, biotechnology, waste water treatment and many industrial applications.

The commercial interest of utilizing chitin and its derivatives lead to the need of inexpensive, reliable source of active and stable chitinase preparations. Current limitation is the instability of enzyme in narrow pH and temperature range [5]. The aim of this study is to optimize chitinase production from locally isolated fungus *Trichoderma virens* UKM-1 in a 2.0 L bioreactor.

* Corresponding Author. E-mail: madihah@bio.fs.utm.my, Tel: +60-7-5534320

2. Material and Methods

2.1 Fungus

Locally isolated fungal strain identified as *Trichoderma virens* UKM-1 was used as chitinase producer. The culture is grown on Potato Dextrose Agar. Upon maturation, spores were harvested with 1.0% (v/v) Tween 80 by lightly scraping the culture surface using a glass hockey stick. Spores were separated by centrifuging at 4000 rpm, 4°C for 30 minutes where the supernatant is discarded. Spores were resuspended with sterile distilled water to spore concentration of 1×10^5 spores/ml. Fungus stock is stored in beads systems at -80°C.

2.2 Fermentation Medium

The composition of *Trichoderma Minimal* medium [6] includes (g/L): colloidal chitin, 20; CaCl₂, 0.3; KH₂PO₄, 3.0 and MgSO₄, 0.3. The pH value of media was adjusted to pH 5.5 with 0.1 M NaOH or 0.1 M HCl prior to sterilization at 121°C, 15kPa for 20 minutes. 10% (v/v) of spore suspension were inoculated into medium. Working volume of 50 ml was used for 250 ml shake flask fermentation while for bioreactor fermentation the working volume is 1.5L.

2.3 Colloidal chitin preparation

20 g of commercial chitin from crab shell (Fluka) was stirred into 200 ml of concentrated HCl. Mixture was continuously stirred with slight heating of 40°C for 5 minutes. Cold distilled water was added in excess amount. Chitin was allowed to precipitate and washed with distilled water until pH reached neutral.

2.4 Assays

Chitinase activity was determined using DNS method [7] with modifications. 1.0 ml of sample supernatant was incubated with equal volume of 10% (w/v) colloidal chitin in 0.2mM phosphate buffer pH 6.5 and incubated at 50°C for one hour. Reaction was terminated by adding 1.0 ml NaOH 0.1M and boiled for 5 minutes. Enzyme suspension is obtained after centrifuging at 4000 rpm for 5 minutes. 1.0 ml of enzyme was mixed with 1.0 ml DNS solution and boiled for 5 minutes. 5ml of distilled water was added to solution and reading was observed at UV absorbance of 535 nm. One unit (U) of chitinase activity is defined as the amount of chitinase needed to catalyze the release of 1µmol of *N*-acetyl-*D*-glucosamine. Protein concentration was assayed using the method of Lowry [8].

2.5 Determination of optimum pH and temperature

The effect of pH on chitinase activity was determined by incubating enzyme at 50°C in varying composition of standard assay buffer between pH 1 to 13 (KCl-HCl buffer, pH 1-2; citrate buffer, pH 3 and 5; citrate-phosphate buffer, pH 3-7; sodium acetate buffer, pH 4-5; phosphate buffer, pH 6-8; KCl-NaOH, pH 12-13). The effect of temperature was determined by incubating enzyme in citrate-phosphate buffer of pH 4 at various incubation temperatures ranging from 30-70°C. Enzyme stability for each parameter was studied by prolonging the incubation period to 2 hours.

3. Results and Discussion

3.1 Shake flask Studies

3.1.1 Optimizing Substrate Concentrations

In this study, colloidal chitin that acts as carbon source was used as substrate. Optimum substrate concentration is crucial to avoid substrate inhibition or fungal cell wall utilization. Fungal cell wall itself contains chitin that may be used for chitinase production [6, 9, 10]. Chitinase can be produced even without the availability of substrate in minimal medium. Infact, chitinase activity is higher in fermentation flaks without colloidal chitin compared to fermentation in flasks with certain colloidal chitin concentrations. This shows that substrate competition occurs between fungal cell wall and colloidal chitin. The omission of nitrogen source in *Trichoderma* minimal medium caused fungal hyphae to be partially empty and broken at the tips. Fragmentation of hyphae releases chitinase into medium [11]. For insoluble substrate such as colloidal chitin, access to enzyme reaction sites on this polymer is restricted to enzyme diffusion. The number of enzyme binding active sites on the substrate exceeds the number of enzyme molecules, unlike in the case of soluble substrates like glucose [12]. The highest chitinase activity was obtained at chitinase activity of 0.109 U/ml in the presence of 4.0% (w/v) of colloidal chitin as presented in Figure 1.

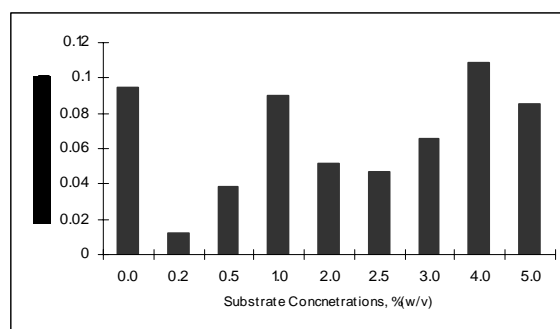


Figure 1: Effect of colloidal chitin concentration on chitinase production

3.1.2 Screening and optimizing stimulant

The complexity of fungal fermentation for chitinase production is caused by mycelia formations which affect the mixing and mass transfer limitations [11] and also the insoluble nature of chitin substrate. Felse, 2000 [13], reported that chitinase production by wild type fungus is much lower compared to bacterial fermentation. Stimulant is required to improve chitinase production as presented in Table 1.

The fermentation medium used in this study consisted of carbon source and several macronutrients such as K^+ , Mg^{2+} and Ca^{2+} . Micronutrients are important in maintaining the structural configuration of enzymes [14]. However, Nawani, 2004 [15] commented that the increase of trace elements would decrease chitinase production. Both T2 and Absidia trace elements contains Zn^{2+} , Mn^{2+} and Fe^{2+} which were essential for microbial growth. Medium supplemented with T2 trace elements showed a slight increase in chitinase production compared to medium containing Absidia trace elements. This is because the presence of these trace elements reduced fragmentations and improves hyphae elongation. It was observed that the hyphae were less fragmented and longer compared to flasks containing Absidia trace elements. Absidia trace elements also contain Cu^{2+} and Co^{2+} . These heavy metal ions were suspected to interfere with the active binding sites of chitinase enzyme.

Methanol had been reported as the best type of alcohol being used as a stimulant to enhance primary and secondary metabolite productions such as citric acid or kojic acid. However, no reports had been found which studied the effect of alcohol in enzyme production. Fungal enzymes were known to be cell wall associated and enzyme activities were mainly found in the cell wall of mycelia in submerged fermentations [16]. Methanol promotes fungal cell wall permeability [17]. Therefore, with known occurrence of mycelia fragmentations, addition of methanol would further stimulate the excretion of chitinase into medium. Optimum methanol concentration is also crucial to avoid toxicity. Results in Figure 2 shows that only a slight addition of methanol would increase chitinase activity, while high methanol concentrations would decrease chitinase activity as high methanol concentrations would disturb fungal metabolism and mycelium morphology [17].

Table 1: Comparison of chitinase activity in different types of stimulants at 0.1% (v/v) concentration

Stimulant (0.1% v/v)	Max Chitinase Activity (U/ml)	Protein Concentration ($\mu\text{g/ml}$)
T2 trace elements	0.119	31.818
Absidia trace elements	0.055	196.364
Methanol	0.146	62.727

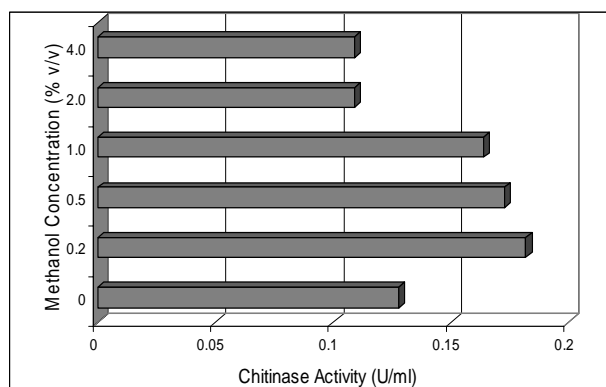


Figure 2: Chitinase activity in different methanol concentrations

3.2 Bioreactor fermentation

3.2.2 Agitation Speed

Further study on chitinase production was investigated in a 2.0L B. Braun Biostat-B bioreactor. Fermentations were carried out at a constant temperature of 30°C with air flow aeration of 1.0 l/m but at different agitation speeds of 200, 250, 300, 400 and 500 rpm. Result on the effect of different agitation speeds is depicted in Figure 3. Agitation play important role in fungal fermentation for mixing, efficient oxygen transfer and heat transfer [18]. In submerged bioreactor fermentation, agitation intensity may affect both fungal morphology and oxygen transfer. Agitation speed also causes morphological differences both in shake flask and bioreactor, resulting in different productivity of target enzyme [19]. Strong agitation would lead to free filaments formation while slow agitation would cause clumps and wall growth. Agitation speed of 250 rpm produces the highest chitinase activity with 0.447 U/ml. It was also observed that low agitation of 200 rpm formed clumps by the third day of fermentation. Wall growth was also observed on both oxygen and pH probes, baffles and impellers. By the fifth day of fermentation, fungus clumped on baffles and impellers. Clumping cause low oxygen transfer leading to poor chitinase production. Not only that, it would deprive fungus from oxygen. However, increase of agitation intensity of more than 250 rpm would not increase chitinase activity. This was due to strong mechanical force caused by impeller that may lead to mycelia damage [16]. In this study, at 300 and 400 rpm, chitinase production decrease steadily. However, at 500 rpm, chitinase production was slightly increased. It was suspected increase of chitinase was not related to chitin consumption. Instead, at this agitation speed, the mechanical force caused mycelia breakage, thus leading to excretion of chitinase by the damaged mycelia. Improvement made by each of optimization process is summarized in Table 2.

Table 2: Improvement in chitinase production after each optimization process

		Chitinase Activity (U/ml)	% of Improvement in chitinase production compared to unoptimized conditions
<i>Shake flask fermentation</i>			
Unoptimized	2.0% (w/v) colloidal chitin	0.052	-
<i>Optimization</i>			
Colloidal Chitin Concentration	4.0% (w/v) colloidal chitin	0.109	52.3
Inducer selection	0.1% (v/v) different types of inducers	0.146	64.4
Methanol Concentration	0.2% (v/v) methanol	0.182	71.4
<i>Bioreactor Study</i>			
Effect of agitation speed	250 rpm	0.447	88.4
	200 rpm	0.076	31.6
	300 rpm	0.109	52.3
	400 rpm	0.064	18.8
	500 rpm	0.146	64.4

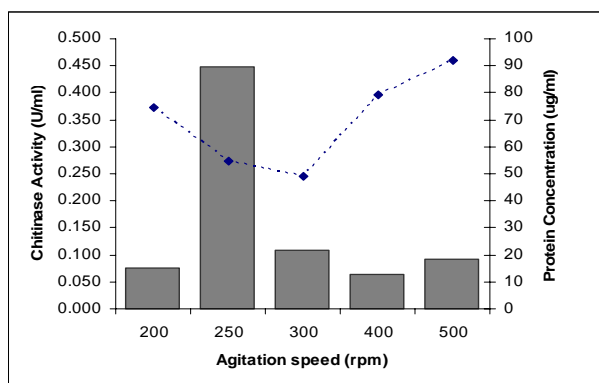


Figure 3: The direct effect of different agitation speed towards chitinase activity and protein concentration

3.3 Characterization of Crude Chitinase Enzyme

3.3.1 Effect of pH upon crude chitinase activity

Crude chitinase activity was found to be the most active citrate phosphate buffer compared to other buffer as shown in Figure 4. It was found that crude chitinase activity was optimum at pH 4 with more than 80% of its relative chitinase activity was retained at acidic condition of pH 3 and pH 5 as in Figure 5. This result was similar to chitinases produced by *Penicillium aculeatum* and *Verticillium lecanii* [1, 2, 20]. At pH value 6, enzyme activity started to decline giving relative chitinase activity of only 18.5% while at pH 7.0 only 3.7%. Eventhough chitinase activity declined, crude chitinase was found to be stable. At pH 3, 100% of its activity was retained and up to 350% of crude chitinase activity was retained at pH 7 as illustrated in Figure 6.

3.3.2 Temperature Optimum and Stability

The activity of crude chitinase was found to be optimum and stable at 45-50°C as presented in Figure 7. This result is similar to purified chitinase of *B. cereus* 6E1 *Chi36* as reported by Wang *et al.*, 2001 [2]. This study also shows that chitinase is unstable at temperature above 50°C as shown in Figure 8. At temperature 70°C, only 37.5% of its residual activity was retained.

4. Conclusion

Chitinase production by *Trichoderma virens* in *Trichoderma* minimal medium at optimum condition is improved by 88.4% in bioreactor compared to unoptimized condition in shake flasks. Optimization of substrate was conducted in shake flask where 4.0% (w/v) of colloidal chitin was the optimum substrate concentration. Other optimization works done in shake flasks include stimulant screening and optimization. It was found methanol was the best stimulant at 0.2% (v/v). Further optimization conducted in a 2.0L Biostat B bioreactor with optimum agitation speed of 250 rpm resulting in 88.4 % improvement. Crude chitinase produced was found to be optimum at pH 4 and at 50°C. Crude chitinase was stable at pH nearing neutral but, crude chitinase stability decreases as temperature increases from 50-70°C.

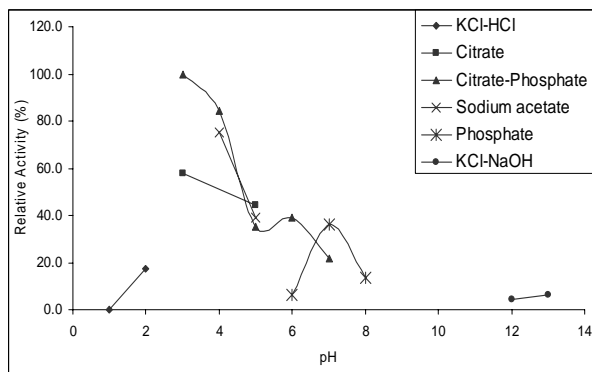


Figure 4: Effect of pH on chitinase activity. Crude chitinase was incubated at 50°C for 1 hour with 10%(w/v) colloidal chitin in KCl-HCl buffer (pH1-2), Citrate buffer (pH 3,5), Citrate-Phosphate buffer (pH 3-7), sodium acetate buffer (pH 4-5), phosphate buffer (pH6-8) and KCl-NaOH (pH 12-13).

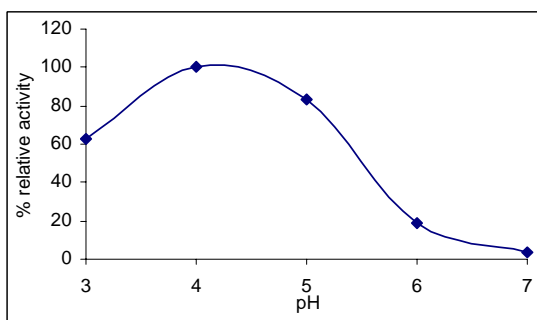


Figure 5: Effect of pH on chitinase activity at 50°C

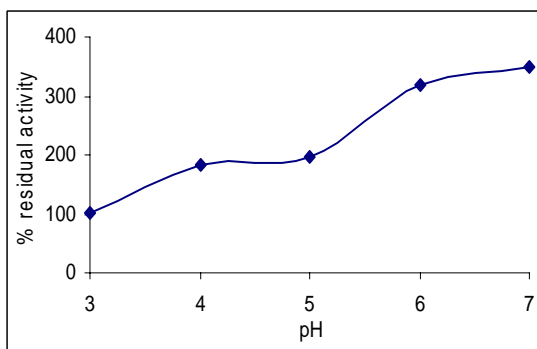


Figure 6: Percentage of residual chitinase activity in various pH at 50°C

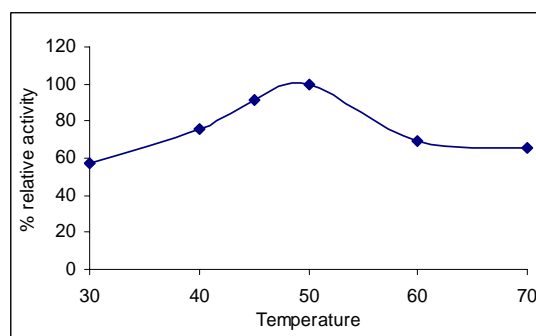


Figure 7: Effect of temperature on chitinase activity at pH 4.

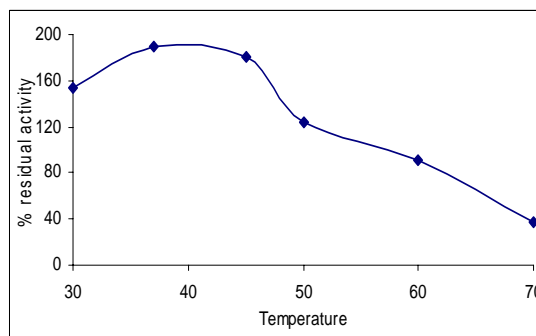


Figure 8: Percentage of chitinase activity in various temperatures at 50°C

References

- [1] Binod, P.; Pusztahelyi, T.; Nagy, V.; Sandhya, C.; Szakács, G.; Pócsi, I.; and Pandey, A. 2005. Production and purification of extracellular chitinase from *Penicillium aculeatum* NRRL 2129 under solid-state fermentation. *Enzyme and Microbial Technology* Article in Press
- [2] Wang, S.L.; Moyne, A.L; Thottappilly, G.; Wu, S.J.; Locy, R.D.; Singh, N.K. 2001. Purification and characterization of a *Bacillus cereus* exochitinase. *Enzyme and Microbial Technology* 28: 492-498
- [3] Gokul, B.; Lee, J.H.; Song, K.B.; Rhee, S.K.; Kim C.H. and Panda T. 2000. Characterization and applications of chitinases from *Trichoderma harzianum*- A Review. *Bioprocess Engineering*. 23. 691-694
- [4] Patil, R.S.; Ghormade, V. and Deshpande M.V. 2000. Chitinolytic enzymes: an exploration. *Enzyme and Microbiology Technology*. 26. 473-483
- [5] Yuli, P. E.; Suhartono, M. T.; Rukayadi, Y.; Hwang, J.K.; Pyun, Y.R. 2004. Characteristics of thermostable chitinase enzymes from Indonesian *Bacillus* sp.13.26. *Enzyme and Microbial Technology* 35: 147-153
- [6] Ulhoa, H.J. and Peberdy, J.F. 1999. Regulation of chitinase synthesis in *Trichoderma harzianum*. *Journal of General Microbiology*. 137. 2163-2169

- [7] Miller, G.L. 1959. Glucose Assay by Dinitrosalicylic Colorimetric Method. *Analytical Chemistry*. 31. 426
- [8] Walker, J.M. ed. 1994. *Basic Protein and Peptide Protocols*. Humana Press.
- [9] Andrade, V.S.; de Barros N.B.; Fukushima, K. and de Campos T.G.M. 2003. Effect of medium components and time of cultivation on chitin production by *Mucor circinelloides* (*Mucor javanicus* IFO 4570) – A factorial study. *Rev Iberoam Micol*. 20. 149-153
- [10] El-Katatny M.; M.Hetta A.; M.Shaban G. and M.El-Komy H. 2003. Improvement of cell wall degrading enzymes production by alginate encapsulated *Trichoderma* spp. *Food Technology Biotechnology*. 41(3). 219-225
- [11] Shuler, M.L., and Kargi F. 2002. *Bioprocess Engineering-Basic Concepts*. 2nd Edition: Prentice Hall
- [12] Papagianni, M. 2004. Fungal morphology and metabolite production in submerged mycelial processes. *Biotechnology Advances* 22: 189-259
- [13] Felse, P.A and Panda, T. 2000. Submerged culture production of chitinase by *Trichoderma harzianum* in stirred tank bioreactors – the influence of agitator speed. *Biochemical Engineering Journal*. 4. 115-120
- [14] Madigan, M.T.; Martinko, J.M. and Parker J. 2000. *Biology of Microorganisms*. 9th Edition: Prentice Hall International
- [15] Nawani, N.N and Kapadnis, B.P. 2004. Optimization of chitinase production using statistics based experimental designs. *Process Biochemistry*. Article in press
- [16] Wang, L.; Ridgway, D.; Gu, T. and Moo-Young M. 2005. Bioprocessing strategies to improve heterologous protein production in filamentous fungal fermentations. *Biotechnology Advances*. 23. 115-129
- [17] Ul-Haq, I.; Ali, S.; Qadeer, M.A. and Iqbal, J. 2003. Stimulatory effect of alcohols (methanol and ethanol) on citric acid productivity by a 2-deoxy-D-glucose resistant culture of *Aspergillus niger* GCB-47. *Bioresource Technology*. 86. 227-233
- [18] Felse, P.A. and Panda, T. 1999. Self-directing optimization of parameter for extracellular chitinase produced by *Trichoderma harzianum* in batch mode. *Process Biochemistry*. 34. 563-566
- [19] Liu, B.L.; Kao, P.M.; Tzeng, Y.M. and Feng, K.C. 2003. Production of chitinase from *Verticillium lecanii* F091 using submerged fermentation. *Enzyme and Microbial Technology* 33: 410-415
- [20] Fenice, M.; Selbmann, L.; Di Giambattista, R. and Federici F. 1998. Chitinolytic activity at low temperature of an Antarctic strain (A3) of *Verticillium lecanii*. 149. 289-300

Degradation of Thiosulfate by Sulfide-Oxidizing Enzyme Produced by Bacteria Locally Isolated from Effective Microorganism Active Solution (EMAS)

Nur Suhanawati Ashaari¹, Madihah Md. Salleh^{1*} and Rosli Md. Illias²

¹ Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia.
Tel: +60-7-5532526, E-mail: suhana_bio@yahoo.com

² Faculty of Chemical and Natural Resources Engineering,
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia.
Tel: +60-7-5535804

Abstract

A potential aerobic sulfur-oxidizing bacterium (SOB) that was believed to be sulfide-oxidizing enzyme producing strain was previously isolated from Effective Microorganism Active Solution (EMAS), currently known as SO₂. This gram negative bacterium was capable of growing autotrophically in sulfur-oxidizer medium, containing 4mM sodium thiosulfate that serves as an energy sources and electron donor. Sulfate ions were the expected end product of sulfide oxidation catalyzed by extracellular sulfide-oxidizing enzyme. The aims of this study are to identify and characterize the SO₂ and to induce the activity of sulfide-oxidizing enzyme in this strain. The activity of sulfide-oxidizing enzyme was determined spectrophotometrically by measuring the increase of sulfate production using BaCl₂ solution, while the oxidation of thiosulfate was colorimetrically determined at 460 nm. One unit of sulfide-oxidizing activity was defined as amount of enzyme required to produce 1 μmol sulfate per min per mL (U). The maximum sulfide-oxidizing activity (0.064 U) was achieved when the strain was grown at pH 5.0, 30°C in medium containing 1% (w/v) peptone as nitrogen sources after 15 hours incubation. The specific growth rate of this strain at this condition was 0.1552 h⁻¹, with doubling time value of 4.47 h.

Key words: sulfur-oxidizing bacteria, thiosulfate, sulfide oxidation

1. Introduction

The removal of hydrogen sulfide (H₂S) from natural or industrial gases is an important concern of environmental technology, since this toxic corrosive gas is extremely hazardous to human health. H₂S is a colorless gas with a strong "rotten egg" smell [1], and their removal is of current interest. It appears naturally as a byproduct of the decomposition process of organic matter or produced in some industries such as pulping site, petroleum refinery plant, drug manufacturing process, sewage treatment facility and livestock raising farm [2].

Conventional physical-chemical processes of H₂S removal are mainly based on oxidation with air or permanganate, precipitation as a metal salt [3] or by direct oxidation by agents like Cl₂, H₂O₂ and NaClO [4]. But, the drawbacks of these methods include the need for specialized facilities, waste disposal, hazardous chemical usage and thus increase the costs of operation due to high energy demand (high pressure and temperature) [5].

Potential biological alternatives for H₂S removal application has been developed by using microorganism that able to oxidize H₂S, producing sulfate or elemental sulfur as a consequence of complete or incomplete metabolism, respectively [4].

A number of microbial processes for H₂S removal have been proposed that are based on oxidation of microorganism that has an ability to oxidize sulfide to sulfate or elemental sulfur. Sulfur-oxidizing bacteria (SOB) are candidates microorganism for accomplishing the removal of H₂S.

Most of the known SOB belongs to the genera *Thiobacillus*, *Thiothrix*, *Thiomicrospira*, *Achromatium* and *Desulfuromonas* [6]. This autotrophic microorganism utilize reduced inorganic sulfur compounds such as sulfide, thiosulfate, sulfite, tetrathionate or elemental sulfur for the biosynthesis of cellular material or transform these compounds as part of a respiratory energy-generating process [6].

However, oxidation of sulfur compounds is not restricted to the true sulfur bacteria; this process also occurs in heterotrophic bacteria isolated from soil and marine environment [6]. Most of the heterotrophic bacteria belong to the genera *Pseudomonas* [7], *Xanthobacter* [8], or are *Escherichia coli* strains [9].

The oxidation of starting substrate such as sulfide, elemental sulfur or thiosulfate will produce sulfate as the end product, with sulfite or polythionates as the intermediate of sulfur compounds oxidation [10-11]. The oxidation of inorganic sulfur compounds was catalyzed by enzyme generally known as sulfide-oxidizing enzyme [12]. These processes of inorganic sulfur compounds oxidation

* Corresponding Author. Email: madihah@bio.fs.utm.my,
Tel: +60-7-5534320

are varying because the mechanism appeared to differ radically among different sulfur bacteria.

In this paper, a potential aerobic SOB isolated from EMAS has been used throughout this study. The scope of this work was to optimize production of sulfide-oxidizing enzyme by manipulating the physical and environmental factors such as pH, temperature and different nitrogen sources to obtain the optimal condition for enzyme production.

2. Materials and Method

2.1 Media and Cultivation Conditions

Sulfur-oxidizer medium for sulfur-oxidizing bacteria contained (per liter)[13]: 10 g of Bacto-Peptone, 1.5 g of K_2HPO_4 , 0.75 g of ferric ammonium citrate and 1.0 g of $Na_2S_2O_3 \cdot 5H_2O$. The pH was adjusted to 7.0 using 1 M HCl before sterilizing by an autoclave. To prepare solid medium, agar was added to a final concentration of 15 g per liter. The SO_2 cultures were incubated at 30°C at 180 rpm shake, and growth was followed by measuring the turbidity at 660nm using the spectrophotometer.

2.2 Preparation of Crude Enzyme Extract

Freshly grown cells were harvested by centrifugation at 4000 rpm, 4 °C for 20 minutes and the supernatant was used as a crude enzyme extract. The crude enzyme was analyzed for sulfide-oxidizing activity, product (sulfate), protein concentration and thiosulfate oxidation.

2.3 Determination of Sulfide-Oxidizing Activity

2.3.1 Reagents Preparation

Sodium sulfide solution 0.02M was prepared by dissolving $Na_2S \cdot xH_2O$ in an alkaline solution consisting 0.16 g of NaOH, 0.02 g of ethylenediaminetetraacetic acid disodium salt ($EDTA Na_2 \cdot 2H_2O$), 2 mL of glycerol and 40 mL of distilled water. The Na_2S solution was prepared just before use.

2.3.2 Sulfide-Oxidizing Assay

The sulfide-oxidizing activity was determined by measuring the product of enzymatic reaction, SO_4^{2-} in the reaction mixture. The enzyme assay protocol was a modified version of Hirano *et al.* (1996) enzyme assay [14].

The reaction was initiated by the addition of 0.5 mL Na_2S solution into the reaction mixture that contain 4.5 mL 0.1 M sodium acetate buffer (pH 5.6) and 1 mL crude enzyme. The mixture was incubated for 30 minutes at 30°C and the reaction was subsequently terminated by addition of 1.5 mL NaOH (1.0 M) followed by thoroughly mixing. The amount of sulfate ion was detected by mixing 2 mL of 10% (w/v) $BaCl_2 \cdot 2H_2O$ and 2 mL of reactant and the absorbance was measured at 450 nm using spectrophotometer [15]. One unit of sulfide-oxidizing activity was defined as amount of enzyme required to produce 1 μ mol sulfate per min per mL (U).

2.4 Analysis of Residual Thiosulfate

The level of thiosulfate in the medium was determined by the colorimetric procedures of Sorbö [16]. This thiosulfate assay reaction occurs at room temperature with the addition of Cu^{2+} ions as catalyst. Thiosulfate was estimated at 460nm after the addition of an acidic solution of ferric nitrate.

2.5 Protein Analysis

Protein was estimated, with crystalline bovine serum albumin as standard, by the method of Lowry *et al.* [17] at 750nm.

2.6 Induction of Sulfide-Oxidizing Enzyme Production

2.6.1 Effect of Various pH

The pH of culture media was adjusted within 4.0 to 9.0 using 1 M HCl or 1 M NaOH. pH adjustment was done before autoclaving process. After autoclaving and cooling, 10% (v/v) inoculum with optical density of 0.6-0.8 was added. The cultures were incubated at 30°C in 180 rpm-shaking incubator and samples were withdrawn at 3 hours interval for the analysis. The optimum pH of culture medium achieved in this step was fixed for subsequent experiments.

2.6.2 Effect of Various Temperatures

The cultures were incubated at various temperatures (30°C, 37°C and 45°C) with pH 5.0. Analysis was performed at 3 hours interval to determine the maximum sulfide-oxidizing activity. The optimum temperature of culture medium achieved in this step was fixed for subsequent experiments.

2.6.3 Effect of Various Nitrogen Sources

Sulfur-oxidizer liquid medium was amended with varying nitrogen sources. Peptone in the original formulation was substituted with other organic nitrogen sources, particularly yeast extract, urea and casein; other inorganic sources, particularly ammonium chloride, ammonium acetate and potassium nitrate. The concentration of nitrogen sources was remained the same as original recipe, at 1% (w/v). Each medium was adjusted to pH 5.0 and the cultures were incubated at 30°C. Sample was withdrawn at regular interval for the analysis.

3.0 Results

Effect of Various pH

The effect of initial pH towards sulfide-oxidizing activity was demonstrated in Table 1. Growth of SO_2 was studied over a wide range of pH, ranging from 4.0 until 9.0, and the maximal sulfide-oxidizing activity was obtained at pH 5.0. The specific growth rate (μ) and doubling time (t_d) values of SO_2 strain in pH 5.0 medium were 0.1552 h^{-1} and 4.47 h, respectively.

The time course of sulfide-oxidizing activity by SO_2 strain at initial pH of 5.0 was presented in Figure 1. The highest sulfide-oxidizing activity (0.064 U) was achieved from culture grown at pH 5.0 with production of biomass,

X_{max} , 0.3625 g/L after 15 hours incubation. At this pH, the specific enzyme activity, 0.0241 U/mg was the highest value as compared to other pHs. A drastic decreased in the enzyme activity was observed when the bacteria was grown in the medium with initial pH at 8.0 and 9.0. However, growth of bacteria was not significantly affected by pH. Growth was varied in the biomass ranged of 0.250-0.4625 g/L at pH between 3.0 – 9.0.

Table 1. Effect of Initial pH towards Sulfide-Oxidizing Activity

pH	μ (h ⁻¹)	X_{max} (g/L)	Activit y (U)	[Protein] (mg/mL)	Specifi c activit y (U/mg)
4	0.1296	0.3875	0.037	3.59	0.0102
5	0.1552	0.3625	0.064	2.68	0.0241
6	0.2415	0.2500	0.045	2.70	0.0167
7	0.3035	0.4125	0.042	2.87	0.0148
8	0.1687	0.4625	0.031	3.34	0.0092
9	0.1004	0.4625	0.024	3.10	0.0077

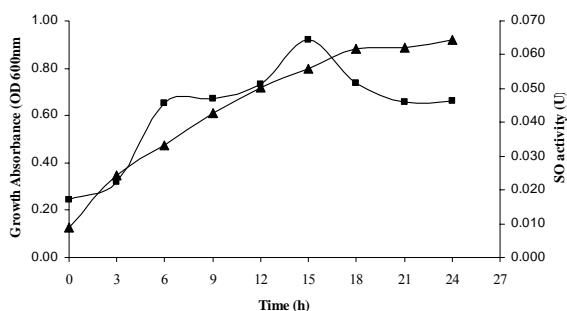


Figure 1. Time Course of Sulfide-Oxidizing Activity (SO Activity) by SO₂ Strain at Initial pH of 5.0. (▲: Bacterial Growth; ■: Sulfide-Oxidizing Activity).

Effect of Various Temperatures

The effect of temperature was studied by incubated the cultures at different temperature such as 30°C, 37°C and 45°C and the result was illustrated in Table 2. From the results, it is clearly shown that microbial growth of SO₂ strain at 30°C resulted the highest sulfide-oxidizing activity. The specific growth rate (μ) and doubling time (t_d) values at 30°C are 0.1552 h⁻¹ and 4.47 h, respectively. The maximum biomass (0.3625 g/L) was produced at this temperature. Temperature higher than 30°C resulted in the decrease of sulfide-oxidizing activity and also the specific growth rate of SO₂ strain. Growth at 45°C was not favorable with a relatively low biomass (0.1500 g/L) and slow doubling time value, 0.1155 h⁻¹.

The time course of sulfide-oxidizing activity by SO₂ strain at 30°C was expressed in Figure 2. The increase of

growth showed a corresponding increase of sulfide-oxidizing activity when cells were in log phase. The maximum sulfide-oxidizing activity (0.052 U) and specific activity (0.0167 U/mg) were attained after 15 hours incubation.

Table 2. Effect of Temperature towards Sulfide-Oxidizing Activity

Temp. (°C)	μ (h ⁻¹)	X_{max} (g/L)	Activity (U)	[Protein] (mg/mL)	Specific activity (U/mg)
30	0.1544	0.3625	0.052	3.09	0.0167
37	0.1477	0.3000	0.042	3.14	0.0135
45	0.1155	0.1500	0.035	3.07	0.0115

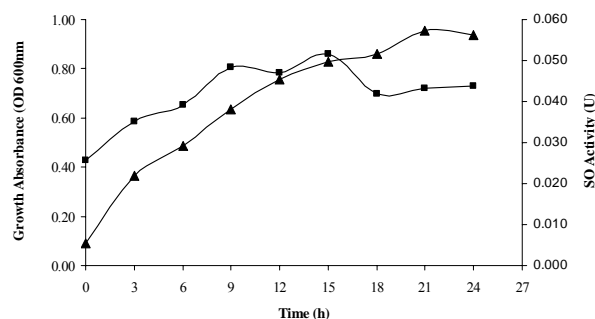


Figure 2. Time Course of Sulfide-Oxidizing Activity (SO Activity) by SO₂ Strain at 30°C (▲: Bacterial Growth; ■: Sulfide-Oxidizing Activity).

Effect of Various Nitrogen Sources

Several kinds of organic and inorganic nitrogen sources were used in these studies as replacement of peptone in the original media formulation. Summarized results for the effect of various nitrogen sources towards sulfide-oxidizing activity are shown in Table 3.

From the experiment, it showed that culture growing in medium supplement with peptone as nitrogen source showed the highest sulfide-oxidizing activity of 0.047 U compared to others, whilst ammonium chloride exhibited the lowest sulfide-oxidizing activity (0.030 U). However, the highest specific activity (0.0437 U/mg) was obtained when urea was used as sole nitrogen source, due to low protein concentration (0.82 mg/mL) produced in this culture. But, growth with urea was not favorable due to poor concentration of biomass produced (0.1375 g/L). The highest biomass level obtained when yeast extract was used in the culture media with value of 1.8625 g/L. Besides that, casein, ammonium chloride and potassium nitrate also promoted good growth of SO₂ strain with their respective specific growth rate, μ were 0.1936 h⁻¹, 0.0889 h⁻¹ and 0.113 h⁻¹.

The time course of sulfide-oxidizing activity by SO₂ strain supplemented with peptone as nitrogen source was demonstrated in Figure 3. Sulfide-oxidizing activity started

to increase as cells were in exponential phase, reaching the maximum level at 15 hours and decrease slowly beyond this point. The amount of protein and specific activity liberated at this time were 3.21 mg/mL and 0.0146 U/mg, respectively.

Table 3. Effect of Various Nitrogen Sources towards Sulfide-Oxidizing Activity

Nitrogen Sources	μ (h^{-1})	X_{max} (g/L)	Activity (U)	[Protein] (mg/mL)	Specific Activity (U/mg)
Peptone	0.1447	0.5125	0.047	3.21	0.0146
Yeast extract	0.1841	1.8625	0.038	1.94	0.0196
Urea	0.0676	0.1375	0.036	0.85	0.0437
Casein	0.1936	1.1125	0.032	3.17	0.0101
Ammonium chloride	0.1011	0.2250	0.030	0.79	0.0433
Ammonium acetate	0.0424	0.1250	0.037	0.98	0.0381
Potassium nitrate	0.1130	0.2250	0.044	0.87	0.0512

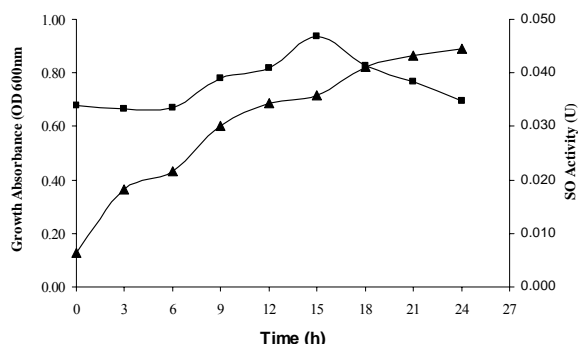


Figure 3. Time Course of Sulfide-Oxidizing Activity (SO Activity) by SO_2 Strain Supplemented with Peptone as Nitrogen Source. (\blacktriangle : Bacterial Growth; \blacksquare : Sulfide-Oxidizing Activity).

Metabolism of Thiosulfate

Thiosulfate is the common oxidizable substrate that is most suitable for investigations of sulfur-oxidizing bacteria processes. In this study, thiosulfate acts as electron donor and energy sources for the sulfur-oxidizing bacteria. The sulfur-oxidizer media was supplemented with 4 mM sodium thiosulfate as substrate for microbial activity. At this concentration, the oxidation of thiosulfate by SO_2 strain was low. The maximum oxidation of thiosulfate that can be achieved at this stage of experiment is when peptone was used as nitrogen source. In medium containing

peptone under optimized condition, the thiosulfate oxidation percentage by SO_2 strain was 6.65% with highest sulfide-oxidizing activity (0.047 U) and considerable amount of biomass (0.5125 U/mg) (Table 3). The results of thiosulfate oxidation by SO_2 strain, at pH 5.0 and 30°C at various nitrogen sources was summarized in Table 4.

Table 4. Percentage of Thiosulfate Oxidation by SO_2 Strain at Various Nitrogen Sources

Nitrogen Sources	Thiosulfate Oxidation (%)
Peptone	6.65
Yeast extract	3.33
Urea	2.30
Casein	2.30
Ammonium chloride	3.33
Ammonium acetate	2.30
Potassium nitrate	6.43

4.0 Discussion

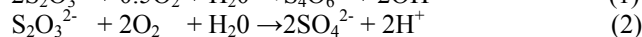
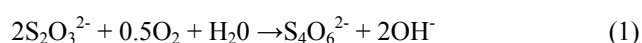
A possible sulfide-oxidizing enzyme that oxidize sulfide to sulfate was detected in the cell-free supernatant of SO_2 strain, based on the enzyme activity obtained. However, the activity of this enzyme was relatively low. Kransil'nikova *et al.*, (2004) also demonstrated low activity of enzyme in his work with value of 0.3 nmol/mg protein/min and 1.2 nmol/mg protein/min for strain N1 and strain SSO that grown in presence of elemental sulfur, respectively [18]. It should be noted that activity of enzymes involved in the metabolism of inorganic sulfur compounds varied depending on type of sulfur compounds added to the medium.

This SO_2 strain use thiosulfate in the growth medium that serves as energy substrate, and electron donor for the oxidation of sulfur compounds. Thiosulfate was used in the growth medium because it is the common oxidizable substrate and more readily soluble in water compared to elemental sulfur.

Induction of sulfide-oxidizing enzyme. Most natural environment has pH values between 5.0 and 9.0, and organisms with optima in this range are most common. Growth of SO_2 strain was observed under acid, neutral and alkaline conditions. This indicated that the SO_2 strain were tolerant to acid and alkaline condition. Optimum pH for the microbial growth in this work was at pH 5.0. From literatures, the optimum pH for sulfur-oxidizing bacteria was varied, depending on the microbial habitat. The haloalkaliphilic *Thioalkalivibrio* isolated from soda lakes growth optimally at pH 10.0 [19] while *Thiobacillus caldus* found in environment such as coal spoil heaps was able to grow down to pH 1.0 [20].

As Table 1 shows that the sulfide-oxidizing activity of SO₂ strain was affected by pH changes. The sulfide-oxidizing activity increase until reaches the optimum pH at 5.0 and start to decrease above the optimum pH. When the pH deviates appreciably from optimum pH, denaturation occurs leading to reduction of catalytic properties.

The metabolic activities of microorganism were followed by changes in pH of their environment as growth proceeds. The pH profile of SO₂ culture grown in medium with pH 5.0 was determined to be constant around pH 7.0-8.0. The pH rises in this culture of sulfur-oxidizing bacteria, might be due to formation of polythionate as products of thiosulfate oxidation rather than the formation of sulfate, according to Reaction 1 [21]. The product of complete oxidation of thiosulfate is sulfate or sulfuric acid, results in lowering of pH and acidification of medium (Reaction 2).



Temperature is one of the most important environmental factors affecting growth and activity of microorganisms. In this work, the optimum temperature for sulfide-oxidizing enzyme production by SO₂ strain was 30°C. Literatures also reported that sulfur-oxidizing bacteria such as *Thiobacillus novellus* and *Thiobacillus ferrooxidans* were found to be grown optimally at 30°C [15, 22].

Sulfide-oxidizing enzyme production was greatly associated with growth, and was found to be affected by microbial growth. Therefore, growth was essential for ensuring a greater sulfide-oxidizing activity. The activity of sulfide-oxidizing enzyme was increase with an increase in temperature, until the optimum temperature was reached at 30°C, and it starts to decrease beyond that temperature. At 45°C, growth started to reduce eventually and activity of enzyme was low. This might be due to denaturation of enzyme that leads to changes in the configuration of enzyme active site and loss of catalytic properties [23]. SO₂ strain was a mesophile microorganism and most protein from mesophiles is inactivated by temperatures above 40°C [23].

Nitrogen is a major element in proteins, nucleic acid and several other constituents in the cell. Nitrogen can be found in nature as both organic and inorganic forms. In the study of various nitrogen sources effect, organic nitrogen sources that were used are peptone, yeast extract, casein and urea while the inorganic nitrogen sources was supplied by ammonium chloride, ammonium acetate and potassium nitrate.

All selected nitrogen sources except urea and ammonium acetate supported the growth of SO₂ strain. Yeast extract and casein produced the higher concentration of biomass, and the specific growth rate was high compared to other nitrogen sources. Yeast extract contain various type of amino acid, while casein contain amino nitrogen which is enough to support the bacterial growth.

Peptone was found to be the best nitrogen sources for sulfide-oxidizing enzyme production, which results in the highest sulfide-oxidizing activity (Table 3) compared to

others. The order of nitrogen sources preferential was peptone > potassium nitrate > yeast extract > ammonium acetate > urea > ammonium chloride > casein. The highest specific activity was obtained in urea-containing medium, due to low protein concentration produced (Table 3). However, urea was not a promising nitrogen sources due to insignificant growth of cells in the medium. In literature, urea has been shown to serve as a nitrogen source for other thiobacilli [24]. Ammonium chloride exhibited the poorest level of sulfide-oxidizing activity. This might be due to inhibitory effect of chloride ions. As reviewed, increased concentration of chloride ions in basal medium would inhibit the oxidation activity [25].

Metabolism of Thiosulfate. Oxidation of thiosulfate with oxygen as electron donor has been demonstrated in SO₂ strain. However, low rate of thiosulfate oxidation was observed with this strain. The SO₂ strain only uses 6.65% (w/v) of thiosulfate. Low thiosulfate oxidation might be due to substrate inhibition at high concentration of thiosulfate.

Conclusion

The microorganism known as SO₂ strain that has been used in this work was able to grow optimally at pH 5.0 and 30°C. The specific growth rate (μ) and doubling time (t_d) for this mesophile microorganism were 0.1552 h⁻¹ and 4.47 h, respectively. The maximum sulfide-oxidizing activity of SO₂ strain (0.064 U) was achieved by employing peptone as nitrogen source under optimized pH and temperature at pH 5.0 and at 30°C, respectively. The highest oxidation rate of thiosulfate has been achieved in the same condition, at initial concentration of thiosulfate, 4 mM with oxidation percentage by SO₂ strain at 6.65%. The highest sulfide-oxidizing activity was obtained during the exponential growth of SO₂ strain, and based on the growth profile, sulfide-oxidizing enzyme could be categorized as growth-associated metabolite or primary metabolite.

Acknowledgments

I gratefully acknowledge Dr Madihah Md. Salleh, Dr Rosli Md Illias, Dr Adibah Yahya and Dr Zainoha for kindly support and share their knowledge in this work.

References:

- [1] Wang, R. 2002. Two's Company, Three's a Crowd: Can H₂S be the Third Endogenous Gaseous Transmitter? *The Federation of American Societies for Experimental Biology Journal* 16: 1792-1798.
- [2] Park, D. H.; Cha, J. M.; Ryu, H. W.; Lee, G. W.; Yu, E. Y.; Rhee, J. I.; Park, J. J.; Kim, S. W.; Lee, I. W.; Joe, Y. I.; Ryu, Y. W.; Hur, B. K.; Park, J. K. and Park, K. 2002. Hydrogen Sulfide Removal Utilizing Immobilized *Thiobacillus sp.* IW with Ca-Alginate Bead. *Biochemical Engineering Journal*. 11: 167-173.

- [3] Gevertz, D.; Jenneman, G. E.; Zimmerman, S. and Stevens, J. 1995. Microbial Oxidation of Soluble Sulfide in Produced Water from the Bakken Sands. In *The Fifth International Conference on Microbial Enhanced Oil Recovery and Related Biotechnology for Solving Environmental Problems*. Oklahoma, 295-309.
- [4] Oprime, M. E. A. G.; Garcia, O. J. and Cardoso, A. A. 2001. Oxidation of H₂S in Acid Solution by *Thiobacillus ferrooxidans* and *Thiobacillus thiooxidans*. *Process Biochemistry* 37: 111-114.
- [5] Sorokin, D. Y. 1994. Use of Microorganism in Protection of Environments from Pollution by Sulfur Compounds. *Microbiology* 63:533-547.
- [6] Das, S. K.; Mishra, A. K.; Tindall, B. J.; Rainey, F. A. and Stackerbrandt, E. 1996. Oxidation of Thiosulfate by a New Bacterium, *Bosea thiooxidans* (strain BI-42) gen. nov., sp. nov.: Analysis of Phylogeny Based on Chemotaxonomy on 16S Ribosomal DNA Sequencing. *International Journal of Systematic Bacteriology* 64(4): 981-987.
- [7] Sorokin, D. Y.; Teske, A.; Robertson, L. A. and Kuenen, J. G. 1999. Anaerobic Oxidation of Thiosulfate to Tetrathionate by Obligately Heterotrophic Bacteria, Belonging to the *Pseudomonas stutzeri* Group. *Federation of European Microbiological Societies Microbiology Ecology* 30:113-123.
- [8] Cho, K. S.; Hirai, M. and Shoda, M. 1992. Degradation of Hydrogen Sulfide by *Xanthomonas* sp. Strain DY44 Isolated from Peat. *Applied and Environmental Microbiology* 58(4):1183-1189.
- [10] Starkey, R. L. 1935. Isolation of some Bacteria which Oxidize Thiosulfate. *Soil Science* 39:197-219.
- [11] Kappler, U., and Dahl, C. 2001. Enzymology and Molecular Biology of Prokaryotic Sulfite Oxidation. *Federation of European Microbiological Societies Microbiology* 203: 1-9.
- [12] Kelly, D. P.; Shergill, J. K.; Lu, W. P. and Wood, A. P. 1997. Oxidative Metabolism of Inorganic Sulfur Compounds by Bacteria. *Antonie van Leeuwenhoek* 71:95-107.
- [13] Visser, J. M.; Robertson, L. A.; Verseveld, H. W. V. and Kuenen, J. G. 1997. Sulfur Production by Obligately Chemolithoautotrophic *Thiobacillus* Species. *Applied and Environmental Microbiology* 63(6): 2300-2305.
- [14] Asku, Z.; Kutsal, T.; Gun, S.; Haciosmanogk, N. and Gholminesad, M. 1991. Investigation of Biosorption of Sodium, Sulfate, Cuprum (II) and Nickel (II) Ions to Activated Sludge Bacteria. *Environmental Technology* 12: 915-921.
- [15] Hirano, T.; Kurosawa, H.; Nakamura, K. and Amano, Y. (1996). Simultaneous Removal of Hydrogen Sulfide and Trimethylamine by a Bacterial Deodorant. *Journal of Fermentation and Bioengineering* 81: 337-342.
- [16] Cha, J. M.; Cha, W. S. and Lee, J. H. 1991. Removal of Organo-Sulphur Odour Compounds by *Thiobacillus novellus* SRM, Sulphur-Oxidizing Microorganisms. *Process Biochemistry*. 34: 659-665.
- [17] Sorbö, B. O. 1957. A Colorimetric Method for the Determination of Thiosulfate. *Biochimica Biophysica Acta* 23: 412-416.
- [18] Lowry, O. H.; Rosenbrough, N. J.; Farr, A. L. and Randall, R. J. 1951. Protein Measurement with Folin Phenol Reagent. *Journal of Biological Chemistry* 193: 265-275.
- [19] Krasil'nikova, E. N.; Bogdanova, T. I.; Zakharchuk, L. M. and Tsaplina, I. A. 2004. Sulfur-Metabolizing Enzymes in Thermoacidophilic Bacteria *Sulfobacillus sibiricus*. *Applied Biochemistry and Microbiology* 40(1):53-56.
- [20] Sorokin, D. Y. and Kuenen, J. G. 2004. Haloalkaliphilic sulfur-oxidizing Bacteria in Soda Lakes. *Federation of European Microbiological Societies Microbiology Reviews*.
- [21] Hallberg, K. B.; Dopson, M. and Lindström, E. B. 1996. Reduced Sulfur Compound Oxidation by *Thiobacillus caldus*. *Journal of Bacteriology* 178(1): 6-11.
- [22] Sorokin, D. Y.; Tourova, T. P. and Muyzer, G. 2005. Oxidation of Thiosulfate to Tetrathionate by an Haloarchaeon Isolated from Hypersaline Habitat. *Extremophiles* 9:501-504.
- [23] Sugio, T.; Domatsu, C.; Munakata, O.; Tano, T. and Imai, K. 1985. Role of Ferric Ion-Reducing System in Sulfur Oxidation of *Thiobacillus ferrooxidans*. *Applied and Environmental Microbiology* 49(6):1401-1406.
- [24] Hogg, S. 2005. *Essential Microbiology*. John Wiley & Sons, Ltd
- [25] Brierley, J. A. and Brierley, C. L. 1968. Urea as a Nitrogen Source for Thiobacilli. *Journal of Bacteriology* 96:573.
- [26] Lu, M. C.; Chen, I. M. and Huang, Y. Y. 2004. Effect of Chloride Ions on the Oxidation of Aniline by Fenton's Reagent. *Journal of Environmental Management* 75:177-182.

Research Challenges on Natural Gas Treating Process

Iwan Ratman^{1,2} and A.F. Ismail^{1,*}

¹ Membrane Research Unit
Faculty of Chemical and Natural Resources Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia

² Brunei Shell LNG Sdn Bhd

Abstract

Natural gas, whether produced from a non-associated or associated gas, contains major impurities such as water vapour and acid gas (H_2S and CO_2). In addition, there are still a number of impurities that potentially can cause problems in downstream processing facilities. These components have an adverse effect on either the process itself or on the product quality.

To meet product specification requirement, natural gas has to be treated either one or several steps. Most of gas treating units applies processes of acid gas removal, gas dehydration and mercury removal. Acid gas removal is one of the major units in the gas treating process to remove H_2S and CO_2 . The technology is always open for development because of wide ranges of influenced factors. Saturated water removal, also known as dehydration process, is mostly considered to avoid any blockage in the process stream when it is applied in the lower temperature environment such as LNG process. The mercury removal process is used to remove mercury content in the feed gas down to levels where it may attack any aluminium-based material in the downstream process. The present technology of the gas treating has not exhausted yet. It is still required more researches to improve current process technology. The main objectives are to improve its performances; to achieve more robust operations and other benefits, such as cost leadership, environmental protection and others implicit/explicit benefits. There are rooms for improvement in the gas treating processes. Research opportunities can be implemented on the acid gas removal technology. They are divided into process design, mechanical design and operational methods to solve frequent operational problems. In the gas dehydration unit, some improvement research can be developed such as mole sieve structure and bed configuration variety. Meanwhile, more research in the mercury removal unit can be focused on sulphur impregnated and bed configuration. The other research topics may be developed, not limited to the topics mentioned on this paper. Thereby, the technology improvement of gas treating will always move on to achieve efficient technology in the future.

Keywords: mixed matrix membrane, gas separation

1. Introduction

Since 3000 BC, Chinese collected natural gas coming out to the surface naturally. Romans, Greeks and others in ancient times used seepages of gas to create ever burning sacred lights for religious purposes. Then after dominie (AD), more references to natural gas through the middle-ages to more recent times. The word of "GAS" was invented by the Flemish scientist Jan Baptista Van Helmont (1577-1644).

In 1790, first liquefaction of gas was founded and natural gas was firstly used in practical way in modern times in Fredonia New York (1821). In 1877, methane as major component of natural gas was firstly liquefied for many purposes. Nowadays, natural gas is one of the energy sources either transported via pipeline or liquefied.

Recent days, the terminology of gas is divided into the following:

- Natural gas : gas that occurs naturally
- Associated gas : gas from reservoirs that also produce crude or condensate
- Solution gas : from reservoirs where it is dissolved in crude oil
- Gas-cap gas : same as solution gas, but including a separate gas-phase

- Non-associated : gas from structures which only can produce gas
 - LNG : Liquefied Natural Gas, stored at - 160 C
 - NGL : Natural Gas Liquids, stored at floating roof tanks
 - LPG : Liquefied Petroleum Gas, stored at ambient pressure or temperature
 - CNG : Compressed Natural Gas. Stored at sphere tanks at certain pressure
 - Dry Gas : Little or no C5+ present
 - Wet Gas : significant C5+ fractions and water are present
- Prior to be utilized, natural gas has to be treated to remove undesired components which may causes problems and to meet its product specifications. Some technology approaches are applied and further improved mainly to efficiently remove gas contaminants. The removal techniques are always developed and challenged to achieve more robust operations and other benefits, such as cost leadership, environmental protection and other implicit and explicit benefits. This paper will describe brief present technology of gas treating processes. Looking at the present operational problems of the present technology gives more challenges of how to improve the treating process as well as to suggest the area of process research for future development.

2. Natural Gas Components

2.1 Hydrocarbons

Most of natural gas either associated or non-associated, contains hydrocarbons, where methane (CH₄) is as major component. Following are typical gas compositions:

Table 1: Hydrocarbons

Component	Chemical formula	Composition, %vol	Atm Boiling point, °C
Methane	CH ₄	> 80	- 161
Ethane	C ₂ H ₆	5	- 88
Propane	C ₃ H ₈	2	- 42
Iso-butane	i -C ₄ H ₁₀	0.4	- 12
Normal-butane	n -C ₄ H ₁₀	0.6	0
Iso-pentane	i -C ₅ H ₁₂	0.3	28
Normal-pentane	n -C ₅ H ₁₂	0.2	36
Hexane	C ₆ H ₁₄	0.3	69
Heptane	C ₇ H ₁₆	0.4	98
Octane	C ₈ H ₁₈	0.4	125

2.2 Gas impurities/ contaminants

Natural gas, whether produced from a non-associated or associated gas, contains major impurities such as water vapour and acid gas (H₂S and CO₂). In addition, there are a number of impurities that potentially can cause problems in downstream processing facilities. These components have an adverse effect on either the process itself or on the product quality. The following table gives an overview:

Table 2: Gas impurities

Impurities	Possible source	Problem	Mitigating measures
Water	- Free from reservoir - Condensed	- Hydrate formation - Corrosion	Hydrate and corrosion management
CO ₂	- Reservoir	- Process blockage - Corrosion	- CO ₂ removal - Corrosion inhibitor
H ₂ S	- Reservoir - Riser (bacteria)	- Process blockage - Corrosion - Catalyst poison	- H ₂ S removal - Corrosion inhibitor
Sand/ dust	Reservoir	- Foaming in Amine system - Blockage	Filters
Mercaptans	Reservoir	- Process blockage - Corrosion - Catalyst poison	Mercaptan removal
Acids	Added to well	- Accumulation in amine unit causing solvent degradation	Water wash upstream amine unit
Nitrogen	Reservoir	- Inert - Sales product off spec	Nitrogen stripper
Mercury	Reservoir	Alumunium material attack	Hg removal
BTX	Reservoir	- Process blockage - Environmental issue	BTX removal
Corrosion inhibitor	Added to minimize	Environmental	Carefull selection

	corrosion		
Glycols	Added to prevent hydrates and reduce corrosion	- Condensate devalued - Propane off spec - Accumulation in amine unit	Removal in process
Methanol	Added to prevent hydrates	- Condensate devalued - Propane off spec - Accumulation in amine unit	Removal in process
PH stabilizer	Added to assist in corrosion and hydrate management	None	

The contents of the above contaminants are varied in the feed gas supply depending on the well characteristic and/or how much chemicals used in the upstream process. The degree of contaminant removal is adjusted to meet product specification and considering aspects of costs (OPEX, CAPEX) and product quality management.

2.3 Typical Gas Product Specification

Nowadays, natural gas is well known being utilized when the growth is faster than oil. In 2020, it is predicted that the world oil and gas demands meet the same point of 115 million barrel oil equivalent per-day (boed). Unlike oil, the most consideration of gas utilization is mostly depending on the factors of customer distance and gas-reserves remote area. The key factor of the gas commercialization is the project economy, which is giving more attractive in gas monetisation. Some other factors will strengthen the project decision.

The gas commercialization is now being configured into several industries, including pipeline gas, LNG, Gas to Liquid (GTL), Independent Power Plant (IPP) and Chemical industries such as Petrochemical etc. The gas product specifications for each industry are varied depending on the Gas Sales Agreement. Nevertheless, a typical specification can be explained as shown on the following table:

Table 3: Typical product specifications

Specification	Pipeline gas and IPP	LNG	Gas to Liquid	Chemical Industry
Quality: Heating value, Btu/SCF Wobbe Index, MJ/m ³	N/A 50, +/- 10%	1050 – 1165 N/A	Unspecifi ed	Unspecifi ed
Composition, %mol Methane Ethane Propane Butane + C ₅ +	Not specified, varied	85-96% 9% max - 2% max 0.1% max	88 – 100 - Min Min Min	90 – 100 Min Min Min Min
Impurities, mg/Nm ³ H ₂ S	5 max 30 max	5 max 30 max	- 0.5 max	- 30 max

Specification	Pipeline gas and IPP	LNG	Gas to Liquid	Chemical Industry
Total S				
Internal Plant requirement	50,000 max	50 max	50,000 max	Non corrosive limit
CO ₂ ppm	Saturated	0.1 max	0.1 max	Saturated
Water, ppm	unspecified	50 max	trace	Unspecified
Hg, ng/Nm ³				

Because of the product specification requirement, the gas treating process is applied to achieve the specified parameter. The treating process technology has been developed since the beginning of the gas industry and being continuously improved until now, mainly for efficiency and effectiveness of the present technology.

3. Gas Treating Process

To meet required gas product specifications, natural gas has to be treated in one or several steps. Most of the discussion is talking about gas sweetening process to remove acid gas, water and mercury in the feed gas. To achieve hydrocarbon specification including heating value, a gas separation technology is applied such as distillation, absorption, etc.

The most attractive topic and always open for development is known as acid gas removal technology because of a wide range of factors influenced to achieve the specification. Saturated water removal, also known as dehydration process, is mostly considered to avoid any blockage in the process stream when it is applied in the lower temperature environment such as LNG process. The mercury removal process is used to remove mercury content in the feed gas down to levels where it may attack any aluminium-based material in the downstream process.

3.1 Acid Gas Removal Process

The level of acid gas concentration in the feed gas is an important consideration for selecting the proper sweetening process. Some processes are applicable for removal of large quantities of acid gas. In some cases, however, processes might not sweeten to product specifications. Other processes have the capacity for removing acid gas constituents to the parts per million (ppm) ranges. Some of the processes, however, are only applicable to low concentrations of acid gas constituents in the feed gas to be treated.

Nowadays, many technologies of acid gas removal have been developed to suit with the required objective. The selectivity of sweetening agent is an indication of the degree of removal that can be obtained for one acid gas constituent as opposed to another. There are sweetening processes which display rather marked selectivity for one acid gas constituent. In other cases there is no selectivity demonstrated and all acid gas constituents will be removed.

There process for which operating conditions can have a marked effect on the selectivity exhibited. Some sweetening agents absorb relatively large amounts of hydrocarbons while others are much more selective for the acid gas constituents.

Occasionally natural gas streams will be sweetened at low pressures. However some processes are unsuitable for removing acid gas under low pressure conditions. Other processes are adversely affected by temperature much above ambient. Some processes lose their economic advantages when large volume of gas is to be treated.

3.1.1 Process Selection

The decision to select a treating process in natural gas application for high CO₂ content gases is shown on the following graph:

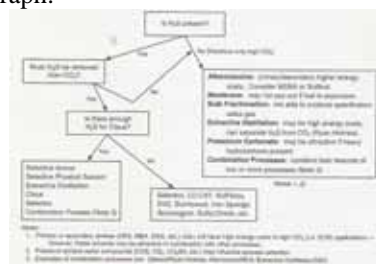


Figure 1: Decision on selecting treating process for high CO₂ content gas

Some general guidance is applied to select the suitable gas treating process. The following graphs illustrate the case of acid gas content in the feed gas. The processes to be considered are dependent on the concentration of CO₂ and H₂S in the inlet gas and on the degree of removal of both constituents required. Graph 1 shows the removal for CO₂ only and Graph 2 for H₂S removal only. The simultaneous removal of CO₂ and H₂S is shown on graph 3. Graph 4 describes a selective removal of H₂S.

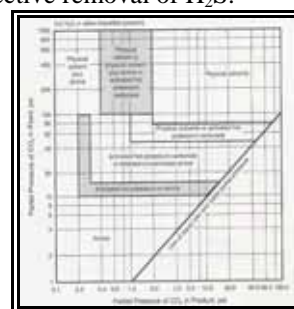


Figure 1: Removal of CO₂ only

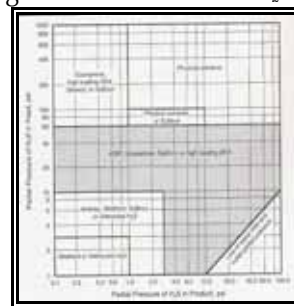
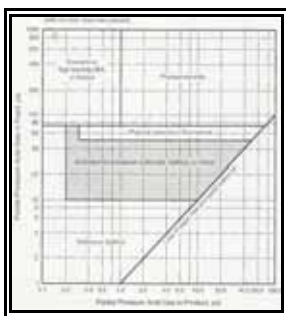
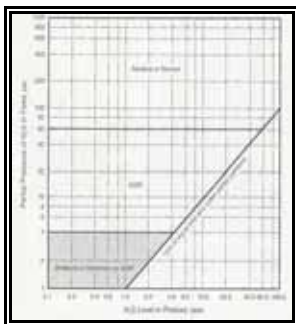


Figure 2: Removal of H₂S only

Figure 3: Simultaneous removal of CO₂ and H₂SFigure 4: Selective removal of H₂S

3.1.2 Frequent Operational Problems

a. Failure to meet treated gas specifications

This problem can be the result of a single factor or many factors working in conjunction. The cause(s) can be identified more quickly and corrected the sooner possible-contributing factors can be ruled out. Monitoring key process parameters that could result in off specification treated gas and responding appropriately before they become a serious problem can sometimes avoid the problem altogether. Following are the potential causes contributing the problem:

- **Feed gas changes:** acid gas concentrations increase, temperature increases/decreases, pressure reduces, and reaches maximum flow (jet flooding).
- **Solvent:** lean solvent too hot, solvent concentration too low/high, inadequate lean loading, circulation rate too low, foaming, degraded solvent and contaminated solvent.
- **Mechanical:** internal damages, feed tray too low in absorber, etc.

b. Foaming

Foaming is the most common cause of upset in the acid gas removal unit, usually resulting in excessive solvent losses, off specification treated gas and a reduction in treating capacity. All of these effects operating cost, revenue and reputation. Operating philosophy with dealing with foaming should be to address the root cause as opposed to continually treating symptoms. Using antifoam agent for temporary relief is acceptable. But for long term, more reliable operations, identifying and correcting the root cause(s) should be taken.

The causes of foaming are well known due to the presence of impurities in the solvent as a result of improper operating

conditions and/or mechanical design. The following impurities can result in severe foaming problems:

- **Hydrocarbon:** with the presence of organic acid can result amine soap.
- **Well-treating fluids:** including acids and corrosion inhibitors
- **Lubricant and welding fluxes:** unclean treatment during maintenance
- **Suspended solids:** including FeS, carbon fines, colloidal iron oxide
- **Raw cotton filters:** poor filter element
- **Make up water:** corrosion inhibitor from steam condensate system and presence of bacteria.

c. Excessive solvent losses

Every acid gas removal unit expect to have some solvent losses associated with its operation. Acceptable losses vary with different application. Best in class category in amine plant, the solvent losses is annually less than 10 percent of total inventory. The average maximum allowable losses in most of units are lower than 30% annually. The excessive solvent losses can have a negative economic impact on operating cost, and it is important to understand how losses can occur and the appropriate corrective actions to take. Losses usually fall into five major categories, which are mechanical, entrainment, solubility, vaporization and degradation of solvents.

d. Corrosion

This problem can result an unscheduled downtime, loss of production, reduction in equipment life, or even injury or death. The corrosion in acid gas removal unit can be controlled and minimize with proper plant design and operating parameters. The key is to minimize acid gas flashing and to understand other contributing factors that can enhance corrosivity of the solvent. The following table describes the type of corrosion and the causes:

Table 4. Type of corrosion

Type of Corrosion	Causes
General corrosion	Uniform material deterioration
Galvanic corrosion	Dissimilar metallurgy coupled together in the presence of an electrolyte
Crevice corrosion	Concentration cells are allowed to form in stagnant areas of the system
Pitting corrosion	Stagnant areas with high halide ion content accelerate the condition
Intergranular corrosion	Most often found in stainless steels in heat affected zones
Selective leaching corrosion	Removal of one element from a metallic alloy
Erosion corrosion	High flow rates and turbulences and presence of solid particles or entrained gas bubbles

Type of Corrosion	Causes
Stress corrosion cracking	Simultaneous tensile stress and corrosive media. Presence of chloride ion in stainless steel
Hydrogen damage corrosion	Presence of hydrogen

3.1.3 Research Challenges

More researches are required to improve process technology of acid gas removal units. The main objectives are to result a better solvent as well as method development to solve problems that are usually encountered in the treating process. Looking at the above frequent problems, the following research challenges can be developed:

Table 5. Acid gas removal – research challenges

Area for Improvement	Challenges	Objective
Process Design	Invention of a new activator to increase kinetic reaction and/or solubility with acid gases (CO ₂ & H ₂ S)	Increase treating capacity
	Improvement on chemical stability against impurities that can lead solvent degradation.	Reduce solvent losses, foaming and erosion corrosion.
	Improvement on combination process on natural gas application	Cheaper technology
	Deep acid gas removal on membrane application	Cheaper technology
	Others than present technologies	Improve performance of present technologies
Mechanical design	Separation unit design to avoid any impurities breakthrough in amine system	Avoid corrosion
	Design improvement on internal contactors	To increase efficiency and reduce solvent losses
Operational Methods	Development of foaming control method	To control foaming
	Improvement on impurities removal in amine system	To avoid corrosion and prevent foaming

Area for Improvement	Challenges	Objective
	Development of operational monitoring tool	To have robust operation

3.2 Natural Gas Dehydration

Dehydration of natural gas is necessary to avoid freeze-ups in extremely cold operation. The water specification for dry gas is typically less than 1 ppm. This application is currently the exclusive preserve of molecular sieves to virtual elimination of other previously used adsorbents like alumina. Although more expensive than alumina, molecular sieve offer advantages of consistently deep dehydration at high capacity under various operating conditions and substantially reduced co-adsorption of heavy hydrocarbons and contaminants leading to vastly improved adsorbent lifetime.

Like acid gas removal process, dehydration process is always developed to achieve mostly robust operation and lifetime prolong. There are currently some alternative processes available in the gas treating industry as follows:

- Glycol dehydration
- Removal of acid gas components involved (IFPEX-1)
- Condensation by compression and/or cooling
- Physical separation through membranes
- Adsorption on activated solid desiccants: silica gel, activated alumina, activated bauxite and molecular sieve.

3.2.1 Process Selection

The general guideline to select appropriate desiccant is shown on the following graph:

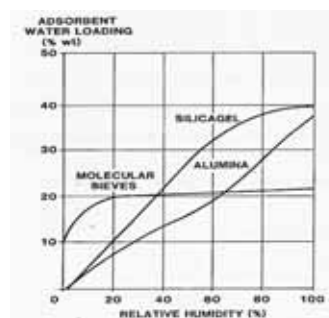


Figure 5: Water adsorbent performance profile

Free and saturated water often accelerates corrosion and ice or hydrates can plug valves, lines and heat exchangers. Molecular sieve, activated alumina and silica gel are used as moisture adsorbent (desiccant). Molecular sieve have a much greater drying capacity at low relative humidity than do either silica gel or alumina, At higher relative humidity (35% and above) silica gel is superior.

When a desiccant must reduce the water content in the process fluid to an extremely low level (less than 1 ppm),

molecular sieves is the desiccant of choice (silica gel and alumina can adsorb water down to 10-100 ppm only). Therefore, nowadays in the natural gas industry, molecular sieves are widely used for dehydration. The following table describes comparisons of desiccants:

Table 6: Comparison of desiccants

Description	Mol sieve	Alumina	Silica Gel
Drying capacity:			
Feed Gas:			
Low humidity	GG	N	G
High humidity	N	G	GG
High temperature	N	BB	B
Product:			
Water content (ppmv)	< 1	10 – 100	10 –
Water dew point (degC)	-85 / -120	-60 / -75	100 - 75
Drying + treating (CO ₂ & H ₂ S)	G	B	B
Co-adsorption of hydrocarbon	G	G	GGG
Acid resistance	B	BB	N
Free water	B	B	BB
Ammonia	B	N	BB
Thermal stability	GG	G	N
COS hydrolysis		Yes, for COS removal	

Note

G : means Good (positive), increasing number of G means increasing its property/value

N : means Neutral

B : means Bad (negative), increasing number of B means increasing its property/value

3.2.2 Bed Configuration

Following graph is typical bed configuration in gas dehydration unit:

Figure 6: Mole sieve Bed configuration

3.2.3 Frequent Operational Problems

a. Hydrates

Hydrocarbons lighter than butane and acid gases such as CO₂ and H₂S for hydrates in presence of free water and certain ranges of temperature and pressure. Hydrates are serious problems in gas processing industry as they can plug valves.

Hydrates are normally eliminated by appropriate process approaches as follows:

- In LNG industry, pre-cooling of feed gas has to be limited not to achieve or lower than hydrate temperature.
- In pipeline gas, a formation of free water is avoided by knocking down in the upstream vessel.
- It is customary to provide methanol injection capability at sensitive points downstream of the drier beds to cover any accidental moisture breakthrough causing hydrate formation.

b. Retrograde condensation

Depending on the natural gas composition, temperature and pressure, retrograde condensation could take place in the mole sieve beds during adsorption. The problem is more likely at higher pressures. The maximum liquid dropout in beds would correspond to the pressure drop between the feed knockout drum and the bed outlet.

The presence of any appreciable amount of liquid hydrocarbon in the gas affects the mole sieve operation in a variety of ways and should be avoided. The adverse effects of retrograde liquids are:

- Liquids wet the mole sieve, the film adding an extra resistance to the mass transfer of water molecules from bulk gas phase to mole sieve active surface
- The wetting of the mole sieves increases their susceptibility to attrition at relatively high bed velocities employed in gas adsorption units, Attrition can further lead to pressure drop problems, causing probably channelling as well.
- Relatively heavy liquid tend to deactivate sieves due to cooking during regeneration

Potential retrograde liquids can be avoided by introducing sufficient heat into feed gas.

c. Liquid (free water) carry-over

Like retrograde condensation, free water carried over into mole sieve beds will potentially damage the mole sieves and contributing to lifetime reduction. A caking problem inside the bed will occur and increase pressure drop drastically (depending on how much free water carried over).

This problem is usually avoided by improving upstream flash vessel operation by modifying internal demister and sometimes to install coalesce device in between flash vessel and the drier beds. This device effectively adsorbs any free water carried over thru the flash vessel. Other than this, maintaining operating temperature is usually done mainly to drop water as much as possible into the upstream vessel.

d. Mechanical failure

Operational experiences is published that mechanical failure of the drier vessel is usually occurred such as failures on bottom support grid, internal probes, Johnson screen, internal wall damage, etc. The most causes of the problem are due to bed caking causing higher-pressure drop, depressurization rate during upset or emergency, etc. Therefore, mechanical strength of the support grid and other internal parts should be in line with the extra pressure drop during emergency situation.

3.2.4 Research Challenges

Although many researches have been developing, some improvements should be focused on capacity increment and stronger binder of zeolite (base mole sieve material). Following are the suggested area for improvement that could be done:

Table 7: Gas dehydration – research challenges

Area for Improvement	Challenges	Objective
Mole sieve structure	Invention of new shape of the molesieve rather than conventional shape of extrudates and beads	Increase water adsorption capacity
	Effects of molesieve particle sizes to mass transfer zone	Increase adsorption capacity and reduce pressure drops
	New stronger of molesieve binder material	Sustain from liquid carry over to avoid caking during operation
	Optimum nominal pore diameter applied in gas dehydration	Minimize hydrocarbon losses
	Particle crushing strength on various binder material	Optimum lifetime
Bed configuration	Optimum bed configuration in various molesieve sizes	Cost saving and minimize operational problem
	The use of silica gel in the top of mole sieve bed for guard layer	Avoid bed caking

Reference:

1. R.N. Maddox, L.F. Sheerar and D.J. Morgan, Gas Conditioning and Processing, Campbell Petroleum Series, (1998) Oklahoma, USA.
2. N.N. Li Ed, “Recent developments in separation science, “ Cleveland: Chemical Rubber Co. (1972), p 135.
3. K.J. Fitzgerald and J.A. Richardson, Hydrocarbon Processing 65, (1966) 110.

Kinetic Modeling of Methane Conversion in the Presence of Co-Feeding to Gasoline Hydrocarbon Products over W/HZSM-5 Catalyst

Kusmiyati^{1,2} and Nor Aishah Saidina Amin^{1*}

¹ Faculty of Chemical and Natural Resources Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5535588, Fax: +60-7-5581463, E-mail: noraishah@fkkksa.utm.my

² Department of Chemical Engineering
Muhammadiyah University of Surakarta, Pabelan, Surakarta, Indonesia
Tel: +62-271-717417, Fax: +62-271-7174174, E-mail: kmy_ums@yahoo.com

Abstract

The kinetic model based on Langmuir-Hinshelwood for the catalytic methane conversion in the presence of co-feeding ethylene and methanol over W/HZSM-5 catalyst was developed. The reaction was carried out in a fixed bed reactor at temperature from 973 to 1073 K and atmospheric pressure. The feed which containing methane + ethylene + methanol + N₂ was reacted in the reactor at different methane concentrations from 50% to 90% v/v and constant ethylene and methanol concentration. The kinetic model reaction mechanism was tested to correlate the experimental data. The correlation between experimental and calculated reaction rate indicates that the model fits the data very well.

Keywords:

Kinetic modeling, methane, co-feeding,
W/HZSM-5, Gasoline

1. Introduction

Extensive efforts have been devoted for the direct conversion of methane to higher hydrocarbons and aromatics. The transformation of methane to higher hydrocarbons and aromatics has been studied under oxidative and non-oxidative conditions [1,2]. In addition, previous works have found the direct conversion of methane to higher hydrocarbons and aromatics in the presence of small amounts of light hydrocarbons [3,4]. In this study, the kinetic of methane in the presence of light hydrocarbons : ethylene and methanol as co-feedings to produce C₅⁺ higher hydrocarbons in gasoline range over W/HZSM-5 catalyst was studied at temperature of 973 to 1073 K.

2. Methods

2.1. Catalyst preparation

The 2 wt. % W/HZSM-5 catalyst was prepared by impregnation method. NH₄ZSM-5 (SiO₂/Al₂O₃=30) which was commercially available (Zeolyst international Co. Ltd.) was converted to HZSM-5 by calcinations at 773 K for 4 h. It was then impregnated with calculated amount of the aqueous

solution of ammonium tungstate (NH₄)₅H₅[H₂(WO₄)₆].H₂O (A.R.). The sample was dried at 383 K overnight and calcined at 383 K for 5 h. The catalyst was crushed and sieved into the size of 35-60 mesh for catalytic testing.

2.2. Catalytic activity

The kinetic measurements of methane conversion in the presence of co-feeding ethylene and methanol to produce higher hydrocarbons in the gasoline range were conducted in a fixed bed differential reactor. Five series of runs were carried out at different temperatures i.e, 973, 998, 1023, 1048, 1073 K. At each temperature, the experiments were carried out using in feed mixtures: methane-ethylene-methanol-nitrogen. The total Gas hourly space velocity (GHSV) of gaseous feed (methane+ethylene+nitrogen) was kept at 1800 ml/g.h. The co-feeding methanol was mixed with gaseous feed at a constant flow rate of 5 ml/h. The concentration of methane was varied from 50 to 90% v/v at a constant ethylene concentration of 5%v/v. The total pressure of the reaction was around 1 atm. Nitrogen was used as an internal standard for calculating the methane conversion and selectivity of the reaction products.. The gaseous reaction products were analyzed on-line by means of a Hewlett-Packard 5890 automated equipped with a thermal conductivity detector (TCD) while the liquid products analyzes were analyzed

* Corresponding Author

using a Flame ionization detector (FID).

2.3. Results and discussion

2.3.1. Effect of temperature

The effect of temperature on methane conversion is presented in Figure 1. The results shown in Figure 1 reveals that methane conversion increase with an increase in temperature from 973 to 1073 K which is in agreement with the results reported by Weckhuysen *et al.* [5]. The methane conversion percentage increased with increasing temperatures was also reported by Shu *et al.* for the methane dehydrocondensation reaction over Re/HMCM-22 at 973, 1023 and 1073 K. However, the formation rates of benzene and naphthalene was decreased at high temperature as 1073 due to deactivation occurring on the Mo/HZSM-5 catalyst [6].

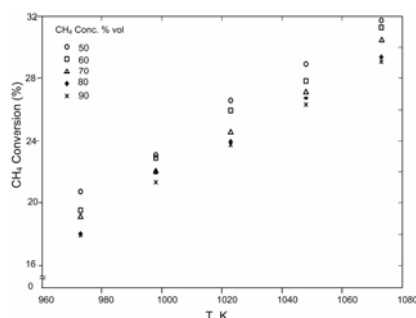


Figure 1. Effect of temperature on the CH_4 conversion

2.3.2. Effect of CH_4 concentration

The influence of methane concentration in the feed consisting of methane-ethylene-methanol-nitrogen at a constant co-feeding ethylene and methanol was also investigated on the conversion of methane to higher hydrocarbons in gasoline range. The result is presented in Figure 2. The result shows that the higher CH_4 conversion was obtained with decreasing methane concentration in the feed. The decreased methane concentration in the feed (more dilution of N_2 in the feed) leads to reactant mixing and the temperature being distributed more evenly along the catalyst bed which affects the methane conversion.

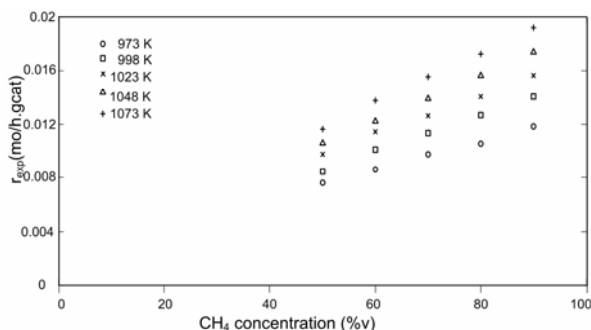


Figure 2. Effect of CH_4 concentration on the reaction rate

2.3.3. Kinetic Modeling

The reactions pathway for the production of C_5^+ hydrocarbons in gasoline range from methane over partially coke deposited on the catalyst (C-W/HZSM-5) is arranged based on Langmuir-Hinshelwood as follows:



where S^* represents the partially coke catalyst=C-W/HZSM-5 catalyst.

The model has been derived based on the above reaction pathway. The assumptions were made to establish the model equation: 1. The surface reaction is the rate-limiting step, 2. C_6H_6 represents the C_5^+ hydrocarbons products in gasoline range, 3. The limitation due to intraparticle diffusion is negligible. Therefore, the rate of methane conversion (r_s) is obtained:

$$r_s = \frac{k_{rs} K_1 P_{CH_4}}{1 + K_1 \frac{P_{CH_4}}{(P_{H_2})^{1/2}} + \frac{K_2}{K_3} P_{C_2H_4}^{1/6} P_{H_2}^{1/2}} \quad (5)$$

$$K_3 = \exp \frac{\Delta G_3}{RT} \quad (6)$$

The equilibrium constant K_3 was calculated using the following relationships for the Gibbs free energy, heat, and entropy for ethylene to benzene equilibrium reaction 4 (eq. 4) [7]:

$$\Delta G_3 = \Delta H_3 - T\Delta S_3 \quad (7)$$

Where:

$$\Delta H_3 = 12296 + 5.456(T - 298) + 1.516 \times 10^{-3}(T^2 - 298^2) - 0.36725 \times 10^{-5}(T^3 - 298^3) \quad (8)$$

$$\Delta S_3 = 0.434 + 5.456 \ln(T / 298) + 3.033 \times 10^{-3}(T - 298) - 0.55087 \times 10^{-5}(T^2 - 298^2) \quad (9)$$

Non linear regression analysis using least squares techniques was done for eq. (10), to determine the unknown parameters i.e, surface rate reaction constant (k_{rs}), adsorption equilibrium constants (K_1 , K_2). The objective function of the residual sum squares was minimizing (eq. 10) to obtain a mathematical fit for the rate reaction equation (Eq. 5):

$$F = \sum_k^n (r_k^{calc} - r_k^{exp})^2 \quad (10)$$

Based on the experiments data, the regression analysis was calculated to obtain the kinetic and equilibrium constants for the reaction of methane in the presence co-feeding methane to produce gasoline hydrocarbons. The results of the comparison between calculated and experimental rate reaction is presented in Figure 3. From the figure, it can be seen that simulated values do not deviate significantly from the experimental value, which indicated that the proposed kinetic model can be used to explain the observed data from methane conversion to C_5^+ hydrocarbons in gasoline range.

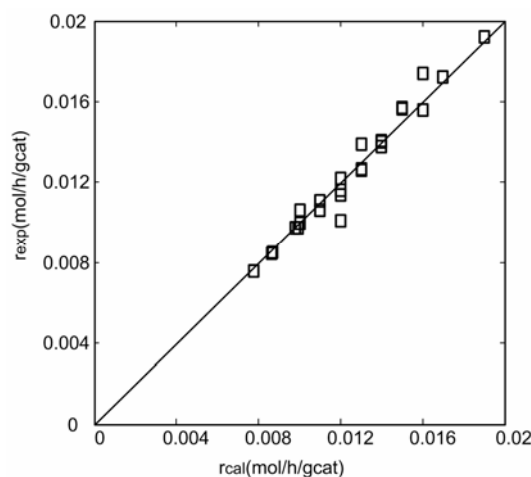


Figure 3. Experimental versus calculated reaction rate

3. Conclusion

Kinetic studies of the methane conversion in the presence of co-feeding ethylene and methanol to produce higher hydrocarbons in gasoline range has been performed over W/HZSM-5 catalyst. The correlation between experimental and calculated reaction rate indicates that the model fits the data very well.

Acknowledgments

The authors would like to express their sincere gratitude to the Ministry of Science, Technology and Innovation, Malaysia (MOSTI) for providing funds to support this research under the Strategic Research project (project no: 09-02-06-0057 or SR 0005/09-07).

Nomenclature

E_a	= activation energy, J/mol
F	= objective function
ΔG	= Gibbs free energy, J/mol
ΔH_i	= heat of reaction i , J/mol
k_0	= frequency factor
k_{rs}	= surface reaction rate constant (controlling step), kmol/kgcat.h atm
K_i	= equilibrium constant
r_s	= reaction rate, kmol/kgmol.h
R	= constant of ideal gas, 8.314 J/mol.K
ΔS	= entropy of reaction i , J/mol.K
T	= temperature, K
W	= mass of catalyst, kg
X_{CH_4}	= methane conversion

Superscripts

calc	= calculated
exp	= experimental

References

- [1] Xu, Y. and Lin, L. 1999. Recent Advances in Methane Dehydroaromatization over Transition Metal Ion-Modified Zeolite Catalysts under Non-Oxidative Conditions. *Applied Catalysis A: General* 188:53-67.
- [2] Kusmiyati and. Nor Aishah, S. A. 2005. Dual Effects of Supported W Catalysts for Dehydroaromatization of methane in the Absence of Oxygen. *Catalysis Letters* 102:69-78.
- [3] Pierella, L. B., Wang, L., Anunziata, O. A. 1997. Methane Direct Conversion to Aromatic Hydrocarbons at Low Reaction Temperature. *Reaction Kinetic and Catalysis Letters* 60:101-106.
- [4] Kusmiyati and Nor Aishah, S. A. 2005. Production of Gasoline Range Hydrocarbons the Presence of ethylene over W/HZSM-5. *Catalysis Today* 106: 271-274.
- [5] Weckhuysen, B.M., Wang, D., Rosynek, M.P. and Lunsford, J.H. 1998. Conversion of Methane to Benzene over Transition Metal ion ZSM-5 Zeolite, I. Catalytic Characterization. *Journal of Catalysis* 175:338-351.
- [6] Shu, Y., Ohnishi R., Ichikawa, M. 2003. Improved Methane Dehydrocondensation reaction on HMCM-22 and HZSM-5 Supported Rhenium and Molybdenum Catalyst. *Applied Catalysis A: General* 252:315-329.
- [7] Iliuta, M.C., Iliuta I., Grandjean B.P.A, and Karachi F. 2003. Kinetic of Methane Nonoxidative aromatization over Ru-Mo/HZSM-5 Catalyst. *Industrial Engineering Chemical Research* 42. 3203-3209.

Isolation, Screening and Characterization of Soluble Exopolymer-Producing Bacteria For Enhanced Oil Recovery

Munirah Tharek¹, *Zaharah Ibrahim¹, S.Hasila Hamzah¹, Noraha Markum¹, Aslizah Mohd Aris¹, Fareh Nunizawati Daud¹, Madihah Md Salleh¹, Adibah Yahya¹, Liew Chong Wai², Nozieana Khairuddin², Rosli Illias², Mohamad Ismail Omar³, Khor Siak Foo³, Ezrin Johanna Elias³, Scott Bailey⁴.

¹ Department of Biology, Faculty of Science
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5532526, Fax: +60-7-5536688, E-mail: munirahtharek@yahoo.com

² Department of Bioprocess, Faculty of Chemical Engineering and Natural Resources
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5535804, Fax: +60-7-5536688, E-mail: r-rosli@utm.my

³ Petronas Research Scientific and Services
Lot 3288 & 3289, Off Jalan Ayer Itam, Kawasan Institusi Bangi, 43000 Kajang, Selangor, Malaysia
Tel: +60-3-89252731, Fax: +60-3-89259702

⁴ Micro Bac International, Inc.
3200 N. IH-35, Round Rock, Texas 78681
Tel: +51-2-3109000, Fax: +51-2-3108800

Abstract

Various types of bacteria isolated from samples of water-oil and palm oil mill effluents (POME) were screened for soluble exopolymers which has potential use in microbial enhanced oil recovery (MEOR) applications. The samples were collected from several Malaysian oil reservoirs and Sedenak palm oil mill. Based on the chemical analyses of the water-oil samples, 6 types of media (HAA, HAG, RAA, RAG, MMS and RGM1) were chosen to initiate bacterial growth. Amongst the 52 pure cultures obtained (44 strains from water-oil samples and 8 strains from POME), 19 strains were stable and able to grow rapidly within 24 hours. These strains were further grown in MM medium in order to enhance exopolymer production. Of these, 5 strains were selected based on the presence of slimy colonies and solubility of the exopolymers produced by the bacteria; T221a exhibited the highest production of soluble exopolymers and was the most heat stable strain. T221a was identified as *Bacillus licheniformis* via biochemical tests and 16S rDNA analyses.

Keywords: Microbial Enhanced Oil Recovery (MEOR), Soluble Exopolymer-Producing Bacteria, *Bacillus licheniformis*.

1. Introduction

The demand of petroleum in the world fuel economy has tremendously increased over the years. However, conventional oil recovery techniques only produced one third of the original oil in place (1). Among the factors affecting the unrecovered oil left in place is differing permeability regions in an oil reservoir. This occurs when

the injected fluid preferentially moves through the more permeable regions and the oil entrapped in the less permeable regions will be bypassed. Therefore, the success of an enhanced oil recovery process ultimately depends on the sweep efficiency, that is, the fraction of the oil reservoir volume contacted by the recovery fluid (2). A method of improving sweep efficiency, thus increasing oil recovery rates is selective plugging (3).

Previous studies had shown that the *in situ* application of the microbial selective plugging process which required microorganisms to grow and produce exopolymers under the environmental conditions that existed in the oil reservoir does result in improving sweep efficiency. Many oil reservoirs are anaerobic and have high temperatures.

* Corresponding Author.

E-mail: zaharah@bio.fs.utm.my,

Tel: +60-7-5534122, Fax: +60-7-5536688

However, little were known about bacteria that could grow and produced polymers under these conditions (2). In this paper, the isolation of bacteria were reported from samples of water-oil and palm oil mill effluents (POME) cultivated at temperatures ranging from 50°C to 100°C in media containing the major components of the aqueous phase of water-oil samples. Since the use of soluble polymers was reportedly successful in enhancing oil recovery (1, 4), therefore in this study only soluble exopolymer-producing bacteria were further screened. The best potential microbe for MEOR application purposes was characterized via biochemical tests and 16S rDNA sequence analyses.

2. Materials and Methods

2.1. Samples

Samples of water-oil and POME were used as microbial sources. The water-oil samples were obtained from selected Malaysian oil reservoirs in Terengganu (Tiong A27, Malong A9, Dulang Western D20L) and Sarawak (Baram G85L, Betty 28L, West Lutong 14L, Alab and Samarang) by Petronas Carigali Sdn. Bhd. Whereas, samples of POME were collected at Sedenak Palm Oil Mill, Johor. *In situ* pH of the samples ranged from 6.0 to 9.0.

2.2. Sample Analyses

Atomic Absorption Spectrometer (AAS) and HACH DR4000 spectrophotometer were used to characterize the components of the aqueous phase of only the water-oil samples to determine the suitable compositions for growth media, thus simulating the reservoirs condition. The NaCl concentrations of the samples were determined using hand refractometer (ATAGO, Type 500).

2.3. Media

Based on the chemical analyses of the formation water, enrichment media used for bacterial isolation were modified mineral salts medium (MMS), *Hyperthermophilic archaea* medium (HAG) (5), modified *Hyperthermophilic archaea* medium (HAA), ravot medium (RAG) (6), modified ravot medium (RAA) and modified reinforced clostridial medium (RGM1). The MMS which was modified from Grula *et al.*, (7) contained the following, in grams per liter: KCl, 0.1; KH₂PO₄, 0.5; CaCl₂·2H₂O, 1.0; NaCl, 18.0; peptone, 2.0; yeast extract, 1.0; glucose, 20.0; MgSO₄·7H₂O, 0.5; K₂HPO₄, 0.5 and NH₄NO₃, 1.0. The RGM1 was modified from Gibbs and Freame (8) with the addition of the following compositions in grams per liter: NH₄NO₃, 2.0 and NaCl, 5.0. The HAA and RAA were modified from HAG and RAG, respectively by replacing glucose with acetate.

Medium for enhancement of the exopolymer production (MM) was modified from Wei (9). MM contained the following compositions in grams per liter: KH₂PO₄, 0.5; K₂HPO₄, 0.5; yeast extract, 3.0; NaCl, 40.0; cysteine HCl, 0.5; NaHCO₃, 10.0; NH₄NO₃, 2.0 and sucrose, 100.0. All media were supplemented with 10 mL per liter (v/v) of filter-sterilized trace element and 10 mL

per liter (v/v) vitamin solution (10). The final pH of media was adjusted to approximately pH 7.0 prior to sterilization at 121°C, 15 kPa for 15 minutes. 2% (w/v) agar was used as a solidifying agent. However, it was replaced by gelrite for cultures incubated at temperatures above 60°C.

Media preparation was based on the desired growth condition. Anaerobic growth was prepared using the Hungate technique where cultures were grown in rubber-stoppered and steel cap serum bottles (11, 12). However, cultures grown on solid media in plates were incubated in anaerobic jars (OXOID). Resazurin solution was added as indicator to anaerobiosis (13).

2.4. Bacterial Isolation

Initial bacterial isolation were carried out anaerobically where 1 mL of samples were inoculated on solid media through pour plate technique and 10% (v/v) of samples were inoculated in 50 mL liquid enrichment media. These cultures were incubated for 5 to 7 days at their respective temperatures which were selected based on their *in situ* reservoir temperature (Table 1). Bacterial survivability was observed via wet mount technique (16), whereas growth in liquid media was measured spectrophotometrically at 660 nm (2). Grown cultures from solid enrichment media were selected and transferred several times on the same media (9). Pure cultures were obtained by serial dilution of liquid enrichment media (14). Techniques used for bacterial isolation on solid media were rolling bottle, shake agar, pour plate, spread plate and dilution streak agar (11,12,13,14,15). Sterile syringe flushed with O₂-free gas was used for sampling of the anaerobic cultures (2). The pure cultures were confirmed based on gram staining of the cells and their colony morphologies. Isolated pure cultures were tested for their stability and ability to grow rapidly within 24 hours. Besides testing the ability of the selected cultures to grow at both anaerobic and aerobic conditions, their optimum incubation temperatures (37°C, 50°C, 60°C, 70°C) were also studied.

2.5. Screening for Soluble Exopolymers

Pure isolates obtained were screened for exopolymer producers based on the ability of isolates to produce slimy colonies on solid media (2, 17). The exopolymer production of selected strains were enhanced in MM medium supplemented with trace element solution (10 mL/L, v/v). Prior to screening for soluble exopolymer producers, exopolymer test was carried out. The test was done according to procedures described in Kim and Fogler (18): cultures (5 mL) were taken and centrifuged (6,000 rpm) for 20 minutes at 4°C to separate soluble and insoluble components. After centrifugation, the supernatant (soluble exopolymers and metabolic byproducts) was kept for the soluble exopolymer test. The pellet collected which consisted of insoluble exopolymers and bacterial cells was rinsed with deionized water before

Table 1: Incubation Temperatures for Bacterial Isolation

Sample	<i>In situ</i> Temperature (°C)	Isolation Temperature (°C)
Dulang Western D20L	93	70-100
Malong A9	104	70-100
Tiong A27	106	70-100
Baram G85L	79	60-80
Betty 28L	89	60-80
West Lutong 14L	77	60-80
Alab	71	50-60
Samarang	64	50-60
POME	45	50-70

centrifuging (6,000 rpm) for 20 minutes at 4°C. After discarding the supernatant, KOH solution (10%, w/v) was added to dissolve the insoluble exopolymer, while keeping the cells intact. The cells and dissolved insoluble exopolymers were separated by centrifugation (6,000 rpm) for 20 minutes at 4°C and the supernatant (insoluble exopolymer dissolved in KOH) was collected for the insoluble exopolymer test. The respective supernatant fractions were treated by adding two volumes of chilled (4°C) ethanol (99.8%) to one volume of supernatant to precipitate out both types of exopolymers (soluble and insoluble) before being left overnight at 4°C. Observations were recorded and samples producing soluble exopolymers were selected.

The soluble exopolymer-producing bacteria were further selected based on the highest exopolymer production and the best heat stable strain. The exopolymer production was observed within 24 hours of incubation at 50°C via product dry weight analyses, whereas heat stability of the strains were determined via thermogravimetric analyses (Perkin Elmer thermogravimetric analyzer model TGA 7) (19, 20). Exopolymer extraction by chilled (4°C) ethanol precipitation was carried out prior to both analyses.

Thermogravimetric analyses was conducted by weighing 10 mg of exopolymer samples. The weighed exopolymers were placed in an analytical balance and furnace was programmed for a linear rise in temperature with time. The samples were weighed continuously with rise of temperature. The temperature was raised at a constant rate of 20°C per minute and heated from 30°C to 900°C. Nitrogen gas was used throughout the analyses to prevent oxidation of the samples. The measurement of weight loss versus temperature was produced on a thermograph. Temperature at material weight loss of approximately 5% was taken as the exopolymer's

3.3. Screening for Soluble Exopolymer Producers

temperature stability (19, 20).

2.6. Strain Characterization

T221a was characterized via biochemical tests which were carried out based on a tentative identification described by Laskin and Lechevalier (21). 16S rDNA sequence analyses were performed to further confirm the characterization result produced via biochemical tests. In this analyses, genomic DNA extraction of T221a, polymerase chain reaction (PCR) amplification of the 16S rDNA and sequencing of PCR products were done as described by Margarita et al. (5) except that universal primers were used for DNA amplification and sequencing reactions. The 16S rDNA sequence was then referred to the gene bank data centre using BLAST system to identify the genus and species of the T221a.

3. Results

3.1. Sample Analyses and Media Selection

Chemical analyses of the formation water were performed to determine major chemical compounds present [22]. However, only the aqueous phase of the water-oil samples was analyzed for the purpose of determining nutrient requirements and trace elements for simulation of the reservoir condition. The most abundant cation was Na⁺ (440-10300 ppm), followed by K⁺ (60-150 ppm). Whereas, NO₃⁻ was found as the most abundant anion which varied from 7 ppm to 17 ppm compared to PO₄³⁻ (0.3-0.6 ppm). As for the NaCl concentrations, it ranged between 14000 ppm to 32000 ppm.

Based on the analyses of the aqueous phase of water-oil samples, 6 different types of enrichment media (HAG, HAA, RAG, RAA, MMS and RGM1) were used to initiate bacterial growth. MM medium which also contained the major components of the formation water was used to enhance exopolymer production of selected pure cultures. Trace elements and vitamin solutions were used to support growth.

3.2. Bacterial isolation

A total of 52 strains of bacteria were isolated as pure cultures; 7 strains from Dulang Western D20L, 4 strains from Malong A9, 25 strains from Tiong A27, 6 strains from Betty 28L, 2 strains from Baram G85L and 8 strains from POME. All strains were isolated under anaerobic condition and survived at their respective temperatures except at 100°C due to Maillard reaction which occurred in media containing glucose. Amongst the isolated bacteria, 20 strains were stable and able to grow rapidly within 24 hours. The selected strain's optimum incubation temperature was at 50°C and showed the ability to grow well at both aerobic and anaerobic condition with optical densities of 0.5 to 2.0 within 24 hours. Consequently, all the forthcoming experiments were carried out at 50°C.

Out of 20 selected strains, 11 strains were determined as exopolymer producers. These strains were then cultivated in MM medium to enhance exopolymer production prior to screening for soluble exopolymer-producing bacteria. Through exopolymer tests (18), 5 strains (TA62bi, T221a, B160, M4B80 and P) were identified producing soluble exopolymers. Amongst the 5 strains, T221a showed the highest heat stability with a weight loss of 5% at 140°C (Table 2) and the highest production of soluble exopolymers (13.57 g/L) (Figure 1).

Table 2: Exopolymers Heat Stability

Strain	Weight loss (%)	Temperature (T°C)
TA62bi	5	112
T221a	5	140
B160	5	76
M4B80	5	100
P	5	120

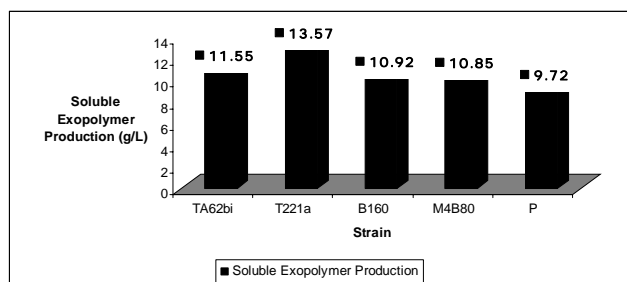


Figure 1: Optimum Soluble Exopolymer Production

3.4. Characterization of T221a

The morphological characteristics of T221a as a gram-positive, motile, rod-shaped, spore-forming bacterium was describe in Bergey's manual as a bacteria that may belong to the genus of *Bacillus* (23). Based on the biochemical tests carried out, T221a showed production of acetoin, enzyme catalase and enzyme gelatinase. T221a also was able to hydrolyse starch and utilize citrate. The positive results in fermentation-oxidation tests determined T221a as a facultative anaerobic bacteria. Following positive results of these tests, ability of T221a to grow in 7% (w/v) NaCl assigned it as *Bacillus licheniformis* (21). Phenotypic characterization via 16S rDNA sequence analyses confirmed T221a as *Bacillus licheniformis* with 98% similarity to the sequence in the gene bank data.

4. Discussion

Enrichment media (HAG, HAA, RAG, RAA, MMS and RGM1) selected for the use of bacterial isolation contained among the major components (Na^+ , K^+ , NO_3^- and PO_4^{3-}) of the water-oil sample's aqueous phase. The presence of Na^+ and K^+ in the media was an advantage due

to its importance in bacterial growth (24). Besides that, the media also contained carbon source, nitrogen source, phosphate source, water and various mineral nutrients which were reported as the basic nutritional requirements for bacterial cultivation (14). These variations of enrichment media were used for the purpose of isolating a larger diversity of bacteria. HAA was selected for the cultivation of *Hyperthermophillic archaea* (5). Whereas, RGM1 had been reported containing compounds suitable for the cultivation of *Clostridium sp* (7). As for RAA and MMS, Wei (9) described them as suitable media for the cultivation of oil reservoir bacteria. Modification of HAA and RAA by replacing glucose as its carbon source with acetate was done to prevent Mailard reaction which affected microbial growth (10). Mailard reaction occurred when media containing glucose were incubated at temperatures higher than 90°C. Agar was replaced by gelrite as a solidifying agent for incubation temperatures more than 60°C because agar's boiling point was at 60°C, whereas gelrite maintained in a solid form at high temperatures up to approximately 100°C (25). MM was used as a exopolymer enhancement medium because it contained all the major components (Na^+ , K^+ , NO_3^- and PO_4^{3-}) of the water-oil sample's aqueous phase in substances such as sucrose and sodium bicarbonate which reportedly enhanced exopolymer production (26, 27, 28).

Of the bacteria isolated, 20 strains were selected amongst the 52 strains based on their ability to grow rapidly within 24 hours and stability of the strains. Stability was addressed by Kim and Fogler (18) as an important factor for MEOR bacteria to sustain growth. This was to ensure that bacteria could survive for a long period of time. The selected strains were also determined as facultative anaerobic bacteria due to its capability to grow well in both aerobic and anaerobic condition. Out of these 20 strains, 11 strains were determined as exopolymer producing bacteria from the presence of slimy colonies on plate (2). These slime-forming bacteria had been reported to have a much greater effect on permeability reduction rather than the non-slimy bacteria (29).

Further screening determined 5 strains as soluble exopolymer-producing bacteria via exopolymer test. Amongst the strains, T221a was selected as the best soluble exopolymer-producing bacteria for MEOR applications due to its highest heat stability and highest production of soluble exopolymers. Heat stability of the soluble exopolymers was an important criteria for *in situ* MEOR applications due to its high reservoir temperature which ranged between 63.9°C and 106.1°C (Table 1). T221a was identified as *Bacillus licheniformis* via biochemical tests and 16S rDNA analyses. The characteristics of T221a as a spore former was an advantage due to the purpose of this isolation study which was to obtain soluble exopolymer-producing bacteria that could survive in the oil reservoirs. This was because spore-forming bacteria were reported to have greater survivability in extreme environmental conditions [30].

5. Conclusion

The information on the composition of the aqueous phase of water-oil samples, pH and the *in situ* reservoir temperatures was very important in selecting the appropriate media and condition to initiate bacterial growth, thus simulating the reservoir condition. Since the reservoir's temperature ranged between 64°C and 106°C, the survivability of cultures at high temperatures were also crucial. Besides that, the stability and ability of the isolated bacteria to grow rapidly within 24 hours were among the criteria of a potential enhanced oil recovery bacteria. Screening for soluble exopolymer producers via exopolymer test was essential. Throughout this study, T221a which was identified as *Bacillus licheniformis* produced the highest production of soluble exopolymers and had the highest heat stability amongst the other strains. Therefore, T221a was selected as the most potential soluble exopolymer producer for microbial selective plugging process in enhancing oil recovery.

Acknowledgments

Authors wish to thank Petronas Research Scientific and Services (PRSS) for samples supplied, Faculty of Chemical Engineering and Natural Resources, Universiti Teknologi Malaysia for the thermogravimetric analyses services and First Base Laboratories Sdn. Bhd., Shah Alam for the DNA sequencing services.

References

- [1] Moses, V., and Springham, D. G. eds. 1982. *Bacteria and the Enhancement of Oil Recovery*. Applied Science Publishers: England.
- [2] Pfiffner, S. M.; McInerney, M. J.; Jenneman, G. E.; and Knapp, R. M. 1986. Isolation of Halotolerant, Thermotolerant, Facultative Polymer-Producing Bacteria and Characterization of the Exopolymer. *Applied and Environmental Microbiology* 51(6): 1224-1229.
- [3] MacLeod, F. A.; Lappin-Scott, H. M.; and Costerton, J. W. 1988. Plugging of a Model Rock System by Using Starved Bacteria. *Applied and Environmental Microbiology* 54 (6): 1365-1372.
- [4] Taylor, K. C.; and Nasr-El-Din, H. A. 1998. Water-Soluble Hydrophobically Associating Polymers for Improved Oil Recovery: A Literature Review. *Journal of Petroleum Science and Engineering* 19:265-280.
- [5] Magarita, L.; Miroshnichenko.; Stackebrandt, H. H. E.; Kostrikina, N. A.; Jeanthon, C.; Nazina, T. N.; Belyaev, S. S.; and Bonch-Osmolovskaya, E. A. 2001. Isolation and Characterization of *Thermococcus sibiricus* sp. nov. from a Western Siberia High-Temperature Oil Reservoir. *Extremophiles* 5: 85-91.
- [6] Ravot, G.; Magot, M.; Fardeau, M. L.; Patel, B. K. C.; Prensier, G.; Egan, A.; Garcia, J. L.; and Olliver, B. 1995. *Thermotoga elfii* sp. nov. A Novel Thermophilic Bacterium from an African Oil-Producing Well. *International Journal of Systematic Bacteriology* 45:308-314.
- [7] Gibbs, B. M.; and Freame, B. 1965. Methods for the Recovery of Clostridia from Foods. *Journal Appl. Bact.* 28: 95
- [8] Grula, M. M.; Russell, H.H.; Janloo, S. M.; and Conway, T. 1991. Effects of Sodium Chloride on Growth and Metabolism on Two Strains of *Clostridium*. *Microbial Enhancement Oil Recovery – Recent Advances* 31:183-206.
- [9] Wei, O, S. 2000. Isolation and Characterization of Indigenous Microorganisms in Malaysian Oil Fields. Master Thesis, Universiti Teknologi Malaysia.
- [10] Cote, R. J.; and Gherna, R. L. eds. 1994. *Nutrition and Media in Methods for General and Molecular Bacteriology (II)*. American Society for Microbiology: Washington D. C.
- [11] Hungate, R. E. 1969. A Roll Tube Method for Cultivation of Strict Anaerobes. *Meth. in Microbiol.* 3B:117-132.
- [12] Balch, W. E.; Fox, G. E.; Magrum, L. J.; Woese, C. R.; and Wolfe, R. S. 1979. Methanogens: Reevaluation of a Unique Biological Group. *Microbiological Reviews* 43:260-296
- [13] Shapton, D.A.; and Board, R. G. 1970. *Isolations of Anaerobes*. Academic Press: London, England.
- [14] Atlas, R. M.; Brown, A. E.; Dobra, K. W.; and Miller, L. 2nd eds. 1988. *Experimental Microbiology: Fundamentals and Applications*. Macmillan Publishing Company: New York.
- [15] Iswaran, V. 2nd eds. 1980. *A Treatise on Media and Methods used in Bacteriological Techniques*. Today and Tomorrow's Printers and Publishers: New Delhi.
- [16] Norell, S. A.; and Messley, K. E. 2nd eds. 2003. *Microbiology Laboratory Manual: Principles and Applications*. Pearson Education, Inc.: New Jersey.
- [17] Smitinont, T.; Tansakul, C.; Tanasupawat, S.; Keeratipibul, S.; Navarini, L.; Bosco, M.; and Cescutti, P. 1999. Exopolysaccharide-Producing Lactic Acid Bacteria Strains from Traditional Thai Fermented Foods: Isolation, Identification and Exopolysaccharide Characterization. *International Journal of Food Microbiology* 51:105-111.
- [18] Kim, D.S.; and Fogler, H.S. 1999. The Effects of Exopolymers on Cell Morphology and Culturability of *Leuconostoc mesenteroides* During Starvation. *Appl. Microbiol. Biotechnol.* 52:839-844.
- [19] Keatch, C. J.; and Dollimore, D. 2nd eds. 1975. *An Introduction to Thermogravimetry*. Heyden and Son Ltd: London.
- [20] Sandrasegaran, S. P. 2004. Selection, Production and Characterization of an Exopolymer from *Cellulomonas Cellulans*. Master Thesis. Universiti Teknologi Malaysia.
- [21] Laskin, A. I.; and Lechevalier, H. A. 1977. *Handbook of Microbiology: Organismic Microbiology Volume 1*. eds. CRC Press: Ohio.
- [22] Bhupathiraju, V. K.; Sharma, P. K.; McInerney, M. J.; Knapp, R. M.; Fowler, K.; and Jenkins, W. 1991. Isolation and Characterization of Novel Halophilic Bacteria from Oil Field Brines. *Microbial Enhancement of Oil Recovery – Recent Advances*

- 31:131-143.
- [23] Benson, H. J. eds. 1990. *Microbial Application: A Laboratory Manual in General Microbiology*. Wm. C. Brown Publishers: United States.
- [24] Gevertz, D.; Paterek, J.R.; Davey, M.E.; and Wood, W.A. 1991. Isolation and Characterization of Anaerobic Halophilic Bacteria from Oil Reservoir Brines. *Microbial Enhancement of Oil recovery – Recent Advances* 31:115-129
- [25] Salleh, M. M.; Ariff, A. B.; Karim, M. I. A.; Khalil, M. S.; Ghani, B. A.; and Harun, A. F. 1997. Isolation of Extremely Thermophilic-Fermentative Bacteria from Petroleum Reservoirs. In *The 3rd Symposium on Trends in Biotechnology & 7th Meeting of MSMBB*. 224-227. Universiti Putra Malaysia.
- [26] Hepburn, I. 1996. Emerging Technology Status Review: Reservoir Biogenics and Its Application to Improved Oil Recovery. The Petroleum Science and Technology Institute.
- [27] Geel-Schutten, G. H. V.; Faber, E. J.; Smit, E.; Bonting, K.; Smith, M. R.; Brink, B.T.; Kamerling, J. P.; Vliegthart, J. F. G.; and Dijkhuizen, L. 1999. Biochemical and Structural Characterization of the Glucan and Fructan Exopolysaccharides Synthesized by the *Lactobacillus reuteri* Wild-Type Strain and by Mutant Strain. *Applied and Environmental Microbiology* 65(7): 3008-3014.
- [28] Chase, K. L.; Bryant, R. S.; Burchfield, T. E.; Bertus, K. M.; and Stepp, A. K. 1991. Investigations of Microbial Mechanisms for oil Mobilization in Porous Media. *Microbial Enhancement of Oil Recovery – Recent Advances* 31:79-94.
- [29] Bryant, R. S. 1987. Potential Uses of Microorganisms in Petroleum Recovery Technology. *Proc. Okla. Acad. Sci.* 67:97-104.
- [30] Bryant, R. S.; and Burchfield, T. E. 1989. Review of Microbial Technology For Improving Oil Recovery. SPE Reservoir Engineering.

The Effectiveness of Environmental Impact Assessment (EIA): A Case Study on Housing Projects

Roslinda Binti Ali¹ and Mansor Bin Ibrahim¹

¹ Kulliyyah of Architecture and Environmental Design
International Islamic University Malaysia, P.O. Box 10, 50728 Kuala Lumpur.
E-mail: roslinda1110@yahoo.com; profmansor@iiu.edu.my

Abstract

A study on the effectiveness of Environmental Impact Assessment (EIA) implementation in housing projects carried out to gauge the followings: 1) to identify environmental problems that arise from housing projects with and without EIA; and 2) to measure the behaviour of residents toward the arisen problems in study areas. A comparison made between housing projects with and without the EIA. Two hundred and thirty-one residents who stayed around the study areas selected as respondents. From the results of the analysis, it concluded that EIA had been effective in minimizing/curbing adverse environmental effects. The respondents from housing area without EIA faced more problems as compared to respondents from housing area with EIA. The problems occurred in housing areas with EIA were temporary in nature and resolved by the time the construction phase end. In contrast to housing areas without EIA, most of the problems were permanent; whereby the problems that existed during construction phase persisted even though the construction had ended. The study also indicated that the non-working group faced more environmental problems compared to the working group as this group spent most of their times at home. The results also revealed that the level of awareness about environmental issues among the public was still low irrespective of their educational background. Therefore, greater efforts should be made to generate awareness. In addition, EIA regulation should be strictly enforced for all housing developments irrespective of size of the area as long as they are likely to have potential negative impact. This study has able to prove that the EIA has been effective and a good planning tool to protect, conserve and preserve the environment.

Keywords

EIA, effectiveness of EIA, comparative study of housing area with and without EIA.

Introduction

Since the enforcement of EIA in Malaysia in April 1988, number of EIA reports received by the Department of Environment (DOE) has been steadily increasing over the years (Ahmad K, 1992). Despite of these, there are still not many studies conducted on EIA in general and on the effectiveness of EIA system in particular (see Ahmad K 1992, Abu Bakar 1993, Dalilah 1993, Jamaluddin 1993, Mohd Ekhwan 1993, Zainudin 1994 & Ahmad Nadzri 1995).

Data gathered from Dewan Bandaraya Kuala Lumpur (DBKL) and Majlis Bandaraya Johor Bahru (MBJB) showed that most of housing developments in Malaysia carried out without conducting EIA study. From the two hundred and forty-two units of housing estates that constructed in year 2000, only twenty-seven units or 11.16% of the total submitted EIA in their construction. In Johor Bahru, on the other hand, in year 2000, of total twenty-eight applications for housing development submitted to MBJB for 'Planning Approval', only four applications (14.29%) submitted have EIA reports. The rest were without EIA reports.

From interviews conducted with the developers, three out of five developers preferred to develop a project without an EIA report. They are of the opinion that conducting an EIA study will only be a burden to them (Zainudin, 1994). These developers still perceived the EIA

as a "stumbling block" to development and only delayed project approval and implementation (Hasmah, 1993: 29). Hence, rather than going through all the hassle, developers prefer the easier way, i.e. develop without the EIA.

Sadly, many failed to see EIA as a planning tool to help them to predict environmental impacts, in addition, to preserve the environment, and from suffering a greater loss if any damage were to occur during construction work. There are many complaints reported in the newspapers over the past years regarding development without a proper EIA or without an EIA at all. In year 2000 only, almost everyday there are complaints in the newspapers regarding impacts on the environment as a result from improper developments. For that reason, this paper is interested to test the effectiveness of approved EIA reports on site, through case studies in the housing projects. To achieve this, a comparison made between housing projects with and without EIA reports.

Research Objectives

The aim of this research is to test the effectiveness of EIA through case studies on the housing projects. In order to achieve this aim, two main objectives formulated as follows: -

- i. To identify environmental problems that arise from both housing projects with and without EIA; and
- ii. To measure the behaviour of surrounding residents toward the arisen problems in both study areas.

Methodology

Two types of data collected for the purpose of the study, i.e. primary and secondary data. Primary data were from site observations, questionnaire distributions and informal interviews with respondents. A set of questionnaire distributed to the residents staying surrounding housing area with and without EIA. Secondary data gathered from the literature review of various sources: books, local and international journals, EIA reports, government publications, seminar papers, internet resources, thesis/ dissertations, local newspapers and databases. Data gathered from field assessed for a comparison between housing projects with and without EIA to see the effectiveness of the approved EIA reports in curbing/minimizing the environmental impacts from happening. These data were analyzed using the Chi Square Test of Independence and the Spearman’s Rank Order Correlation.

The research carried out in the State of Johor Darul Takzim, specifically in Johor Bahru (two housing areas with EIA i.e. Taman Setia Indah and Taman Mutiara Rini) and Segamat (two housing areas without EIA i.e. Taman Mida and Taman Perkasa). Housing projects with EIA selected based on discussion and suggestions from the officer-in-charge at the DOE of Johor. Housing projects without EIA on the other hand, selected based on similar criteria as housing projects with EIA. This is to ensure that the comparison made is based on similar nature and background. All four researched areas located on low land and the housing projects chosen completed within the past three to five years. Since most of the housing projects without EIA were in suburban, Segamat was chosen as the project site for this purpose. This is not because Johor Bahru does not have housing projects without EIA, but the choices were very limited and there was no similarity found between housing projects with and without EIA in Johor Bahru.

Results and Discussion

The questionnaire divided into two parts: part one dealt with the demographic profile of respondents and part two concerned with the environmental problems faced by the respondents because of the housing development surrounding them. However, this discussion will concentrate on three main aspects of findings: a) environmental problems faced by respondents; b) health problems faced by respondents; and c) actions taken by respondents toward the problems.

Environmental problems

Environmental problems here defined as in the forms of air, water, noise and solid waste disposal. List of problems prepared were as follows: 1) open burning; 2) haze; 3) dust; 4) movements of heavy vehicles; 5) noise from construction site after 7p.m.; 6) noise from construction site during weekends or public holidays; 7) damaged roads; 8) traffic congestions due to movements of heavy vehicles; and 9) disposal of garbage and solid waste. A chi square test performed at the start to test the difference between environmental problems that arose during and after construction in housing area with and without EIA. The results shown that there is a significant difference between the numbers of environmental problems arose during and after construction in housing areas that implement EIA and the one that did not implement EIA. When the data further tested using the Spearman’s Rho correlation, the results indicated that there is a significant correlation between environmental problems that arose during and after construction in housing area with and without EIA. The results indicated that the correlation for housing area without EIA (0.968) is stronger than housing area with EIA (0.182). This showed that respondents from housing area without EIA faced more environmental problems than respondents from housing area with EIA.

Environmental problems faced by respondents were also correlated with respondent’s job status to determine the groups’ opinion on the severity of environmental problems. The results showed that the working group (0.811) complained suffering from the problems more frequently as compared to the non-working group (0.730) despite of the fact that the working group spent most of their time away from home. This result perhaps, due to Type II error in analysis, i.e. accepting null hypothesis that there is no difference but in reality there is a difference (Black, 1999: 136). There is quite a big difference between the two sample sizes. The sample size for ‘with EIA’ was twice the size of that for ‘without EIA’. Thus, although many respondents from housing area without EIA complained about facing the environmental problems, the sample size was too small to represent the actual result.

Subsequently, environmental problems faced by respondents according to their job status correlated with types of housing are. This is to verify whether the respondents’ job status influence them in complaining about the arising environmental problems or not. The results shown that housing area without EIA faced more environmental problems as compared to housing area with EIA. It is also interesting to note that it is not influenced by respondent’s status of job. Table 1 summarizes these results.

Table 1: Summary of results for environmental problems faced by respondents

Situation: -	Test: -	p-value: -	Decision rule: -	Notes: -
Environmental problems				
1. The difference between problems during and after				

Situation: -	Test: -	p-value: -	Decision rule: -	Notes: -
construction in housing area with and without EIA. a) During construction b) After construction 2. The relationship between problems during and after construction in housing area with and without EIA a) Housing area with EIA b) Housing area without EIA	Chi square Chi square Spearman's ρ Spearman's ρ	0.00 0.00 0.182* 0.968**	H ₀ rejected H ₀ rejected H ₀ rejected H ₀ rejected	Housing areas without EIA is stronger.
<u>Environmental problems and job status</u> 1. The difference between problems according to job status during and after construction a) During construction b) After construction 2. The relationship between problems during and after construction according to job status a) Working b) Non-working	Chi square Chi square Spearman's ρ Spearman's ρ	0.00 0.182 0.811** 0.730**	H ₀ rejected H ₀ accepted H ₀ rejected H ₀ rejected	Working respondents is stronger
<u>Environmental problems and job status – housing area</u> 1. The relationship between problems during and after construction according to job status in housing area with and without EIA a) Working - with EIA b) Non-working-with EIA c) Working-without EIA d) Non-working-without EIA	Spearman's ρ Spearman's ρ Spearman's ρ Spearman's ρ	0.00 0.241 0.998** 0.881**	H ₀ accepted H ₀ accepted H ₀ rejected H ₀ rejected	Housing areas without EIA are more problematic irrespective of job status

Health problems

To evaluate how far the health problems affected the respondents during and after construction, a Spearman's rho test was performed. The results shown that there is a correlation between health problems that arise during and after construction. Further analysis indicated that respondents in housing area without EIA (0.982) faced more health problems than respondents in housing area with EIA (0.911).

As discussed earlier, the same results obtained when the Spearman's rho correlation tested on health problems. The results indicated that respondents in housing area without EIA faced more health and environmental problems than respondents in housing area with EIA. Not only that, the findings also revealed that the non-working group complained about suffering from health and environmental problems more frequently than the working group. This happened as the non-working group spent most of their time in their houses. However, as in earlier case, it is also important to note that the researcher had also committed Type II error. A summary of results presented in Table 2.

Table 2: Summary of results for health problems faced by respondents

Situation: -	Test: -	p-value: -	Decision rule: -	Notes: -
<u>Health problems</u> 1. The relationship between health problems that arise during and after construction 2. The relationship in health problems during and after construction in housing area with and without EIA a) Housing area with EIA b) Housing area without EIA	Spearman's ρ Spearman's ρ Spearman's ρ	0.959** 0.911** 0.982**	H ₀ rejected H ₀ rejected H ₀ rejected	Housing areas without EIA is stronger
<u>Health and environmental problems</u> 1. The relationship between health and problems during and after construction a) During construction b) After construction	Spearman's ρ Spearman's ρ	0.447** 0.584**	H ₀ rejected H ₀ rejected	Correlation not strong

Situation: -	Test: -	p-value: -	Decision rule: -	Notes: -
<p>Health and environmental problems – job status</p> <p>1. The relationship between health and problems according to job status during and after construction</p> <p>a) Working - during construction</p> <p>b) Nonworking- during construction</p> <p>c) Working - after construction</p> <p>d) Non-working - after construction</p>	<p>Spearman's ρ</p> <p>Spearman's ρ</p> <p>Spearman's ρ</p> <p>Spearman's ρ</p>	<p>0.256**</p> <p>0.577**</p> <p>0.415**</p> <p>0.750**</p>	<p>H_0 rejected</p> <p>H_0 rejected</p> <p>H_0 rejected</p> <p>H_0 rejected</p>	<p>Non-working is stronger</p>
<p>Health and environmental problems – housing area</p> <p>1. The relationship between health and problems during and after construction in housing areas with and without EIA</p> <p>a) During construction- with EIA</p> <p>b) During construction- without EIA</p> <p>c) After construction- with EIA</p> <p>d) After construction- without EIA</p>	<p>Spearman's ρ</p> <p>Spearman's ρ</p> <p>Spearman's ρ</p> <p>Spearman's ρ</p>	<p>0.192*</p> <p>0.314*</p> <p>0.200*</p> <p>0.131</p>	<p>H_0 rejected</p> <p>H_0 rejected</p> <p>H_0 rejected</p> <p>H_0 accepted</p>	<p>Problems in housing areas without EIA are temporary.</p>

Actions taken toward the problems

A list of possible actions also prepared for the respondents to answer to justify their level of awareness in handling matters that related to the environment. Apparently, all the respondents who answered this question chose the option 'no action taken'. When asked for the reasons why they 'do nothing', the answer varied among the respondents who come from a higher and lower educational background (Figure 1). Majority of the respondents who came from educational background of SPM/O-Levels or equivalent and lower, i.e. PMR/SRP or equivalent, primary education and no education at all chose the option, 'do not know where to lodge'. For the option, 'no benefits gained from lodging', most of the respondents came from educational background of SPM/O-Levels or

equivalent and higher, i.e. diploma or equivalent, degree or equivalent and a few with masters and PhD qualifications.

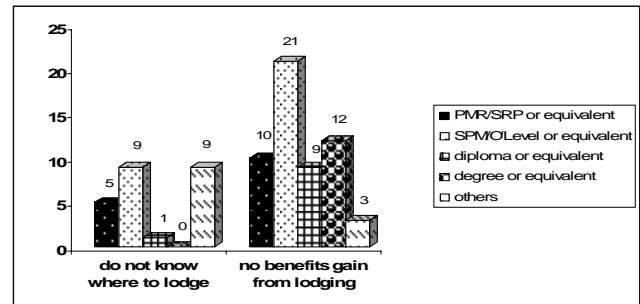


Figure 1: Reasons for not lodging a complaint in relation to levels of education.

It can be concluded that education plays important roles in ensuring that the public know their rights and responsibilities in conserving the environment. But alas, at the same time, a person's educational background is still not enough to generate awareness on environmental issues.

Conclusions

In general, this study has answered the research objectives mentioned above. From this study, it can be concluded that despite of the fact that previous researches had shown that the EIA system implemented in Malaysia is considered as ineffective, the results obtained from this study testified that the EIA reports approved by the DOE proven to be effective managed to curb environmental impacts from occurring. Nevertheless, it could not be denied that there are still problems that arise in housing projects with EIA particularly during construction phase. Although the arisen problems were on temporary basis, this must be resolved so that the problems do not occur in the future. A good EIA report should be flawless and free from any environmental problems that can disrupt a person's health and lifestyle.

Acknowledgements

The authors would like to express many thanks and gratitude to the MOSTI for the grant and IIUM for giving the opportunity to complete and present this research in this seminar. To family and friends, thank you for all the patience, support and encouragement given.

References

Abu Bakar Jaafar. 1993. Environmental Quality Act of 1974: Impact on Design and Development. Paper presented at the PAM-CAA International Conference 1993 on "Design and Development for a Sustainable Future", 3 June 1993, Kuala Lumpur.

Ahmad K Che Ibrahim. 1992. An Analysis of Quality Control in the Malaysian Environmental Impact

Assessment (EIA) Process. Master Dissertation, Dept. of Environmental Biology, Univ. of Manchester.

Ahmad Nadzri Jai. 1995. Kajian EIA: Impak Semasa Fasa Pembinaan dan Pengurusannya di Selangor. Master Dissertation, Universiti Kebangsaan Malaysia.

Black, T.R. 1999. *Doing quantitative research in the social sciences: an integrated approach to research design, measurement and statistics*. London: SAGE Publications.

Dalilah Dali. 1993. Environmental Impact Assessment: Legal Requirements in Malaysia. *Paper presented at the Training Course on Understanding the Environmental Impact Assessment Process*. 12-13 April 1993, Kuantan, Pahang.

Hasmah Harun. 1993. EIA: Malaysia's Viewpoint. *The Surveyor*. 28(1):27-32.

Jamaluddin Md. Jahi. 1993. Impak Pembangunan Selepas Merdeka Terhadap Alam Sekitar dan Pengurusannya Di Semenanjung Malaysia dengan Tumpuan Khas kepada Wilayah Lembah Kelang. PhD Dissertation, Universiti Kebangsaan Malaysia.

Mohd. Ekhwan Toriman. 1993. Keberkesanan Pengurusan Alam Sekitar oleh Pihak Berkuasa Tempatan (PBT) di Negeri Johor. Latihan Ilmiah, Universiti Kebangsaan Malaysia.

Zainudin Zainal Abidin. 1995. Pentadbiran EIA di Malaysia: Analisis Pelaksanaan dan Penguatkuasaan Seksyen 34A, Akta Kualiti Alam Sekeliling 1974 (pindaan 1985) dengan Tumpuan di Selangor dan Wilayah Persekutuan Kuala Lumpur. Master Dissertation, Universiti Kebangsaan Malaysia.

Thermal Ventilation Performance of Malaysia's Single Storey Terrace House

Agung Murti Nugroho¹, Mohd Hamdan Ahmad²

¹ PhD Candidate, Faculty of Built Environment
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-127374703, E-mail: sasimurti@yahoo.co.id

² Associate Professor, Faculty of Built Environment
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5530611, Fax: +60-7-5566155, E-mail: b-hamdan@utm.my

Abstract

Climate is an important factor that can affect not only the physical aspects of human being but also their habits and activity styles. The climatic conditions of the tropical regions are characterized by high air temperatures, high relative humidity and very low wind speeds, which make the environmental conditions uncomfortable. In terms of housing, passive cooling design in tropical climatic regions present the most difficult problem to solve. Terrace houses are a typical example where this problem assumes particular importance. This situation arises from the low air velocity and high temperature observed during day time. Recently, natural ventilation has been increasingly used as an energy-efficient means to provide thermal comfort as well as a healthy indoor environment. This paper discusses the effectiveness of natural ventilation in providing thermal cooling or natural ventilation of a single storey terrace houses in Malaysia. The thermal ventilation performances were investigated using field measurement and Computational Fluid Dynamic (CFD) technique. Validation of CFD FloVent was done by comparing the field measurement. Comparison of the results of simulations and measurements illustrate a good agreement. The simulations were performed on selected day in one year climate condition. The results indicated the single storey terrace house design is not effective in providing natural ventilation for achieving thermal comfort. Further modification of natural ventilation mechanism design must be made and evaluated.

Keywords: thermal ventilation performance, terrace house, natural ventilation

1. Introduction

The tropical humid climate, especially in Malaysia which is mainly characterized by an elevated temperature and high relative air humidity, air flows in building is theoretically required to restore comfort. This emphasises the buildings need to design to provide good ventilation. As found by Fanger and Toftum, the air temperature and humidity combined in the enthalpy have a strong impact on perceived air quality, and perceived air quality determines the required ventilation in ventilation standard [1].

Thermal comfort can be defined as the condition of mind, which expresses satisfaction with thermal environment [2]. To provide indoor thermal comfort, air movement inside naturally ventilated building should be at 0.1–1.5 m/s [3]. A recent study by Griefahn et al. showed that persons were less sensitive to draught when working at an increased activity level than when working at lower activity level. [4]. It is reasonable to consider that all design variables are interdependent. This paper describes the integration of tool in assessing thermal comfort assessment of a single storey terrace house in Malaysia. In this study, the Thermal Data Logger (TDG) was used to determine the temperature variations and Computational Fluid Dynamic (CFD) program to account indoor airspeed for obtaining

thermal comfort (TC) at the height of 1.1 m above the floor for activity while seated (1 met).

2. Methods

2.1. Thermal Comfort Measurement

The main criteria for thermal comfort influence human body as a whole can be divided to environmental parameters: ambient temperature (T_a), mean radiant temperature (MRT), water vapour pressure (Pa) or relative humidity (%), relative air velocity (V_t), and personal parameters: clothing or thermal resistance and activity or metabolic rate (M) [2]. However in this study only air temperature, humidity and air velocity are being used to determine the thermal comfort. The comfort level was measured at 1.10 m above the floor where indoor temperatures and humidity were observed on three rooms (living room, bedroom 1 and kitchen). Completely thermal comfort measurements were obtained on one room (master bedroom) of single storey building (Fig.1a). The thermal data logger measurement was aimed to collect data on environmental parameters (air temperature, mean radiant temperature, relative humidity and wind speed), which were necessary for further thermal comfort analysis. The Dickson data loggers measured dry bulb temperature and

relative humidity, which carefully calibrated with thermal data logger before hand to give reliable results. All measurements (sensor) probe to be placed in the center of the room (Fig. 1b). The metabolic heat production were assumed to be at light activity while seated (1 met) and the clothing insulation was taken from [3] as the characteristic of tropical humid regions. The clothing insulation value was 0.55 clo which is a suitable value for tropical clothing.



Figure 1a (right), 1b (left). A Field Measurement of Terrace House

2.2. Simulation

The simulation of tested model followed field measurement study. All of the tests on five rooms were carried out by making use of CFD Flovent. The basic model in the CFD boundary condition is shown in Fig. 2. The CFD program requires inputs representing problem type, flow domain (material, type of flow, etc.), boundary conditions (walls, inlet, outlet, symmetric wall), and calculation method. The problem type is used to activate calculation modules; in this case, flow, heat transfer, and turbulence modules. The flow domain was set as air density: 1.149 kg/m³, viscosity: 1.872E-005, specific heat: 1007 J/kg.K, conductivity: 0.02643 W/m.K, temperature: 27.6°C-32.2°C. The flow is time independent (steady), subjected to gravity of 9.8 m/s², and is turbulent. Roughness Height (RH) of the ground was set to 0.03. Initial wind was defined at the inlet using an atmospheric boundary layer to obtain a realistic wind profile based on the exponent law (5). The wind speed at 2 m above ground is 0-0.2 m/s based on site Globe Weather data. The turbulent kinetic energy and energy dissipation rate were set to 0.33 and 0.48 respectively. The calculation used the standard k-e epsilon turbulent model with 1500 iterations. To avoid convergence problems, fan relaxation 1 was applied. The solar radiation calculation was set within following limitations (latitude: 1.8 degree, day: 21march, solar times: 8 am-19 pm, solar intensity: 49-210 w/m²).

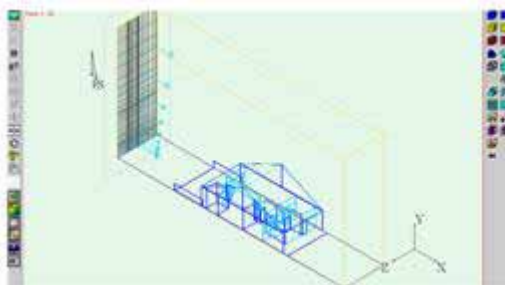


Figure 2. Terrace House Model in CFD

3. Results and Discussion

3.1. Air Temperature and Humidity in Terrace House

Fig. 3 illustrates the temperature inside and outside of the terrace house 12 h during the hottest month and on 21 March 2006. Analysis of the graphs shown in Fig. 3 indicates that the calculated temperature in master bedroom to be higher than other room measurements. One can see from the figure that the daytime outside air temperature could be as highest as 32°C at 12:00. In the morning and in the noon hours, the air inside the building is same than the ambient. Later during the noon, the inside air temperature passes the ambient. One can clearly see that about 15.00 h – 19.00 h, the air at 1.1 m inside the enclosed space is 2–3° hotter than the ambient.

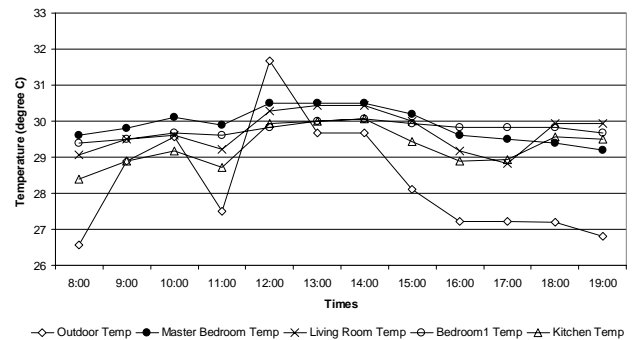


Figure 3. The Temperature Result of Measurements

The ambient humidity is lower to overall room in the house. However, the humidity inside the enclosed space behaves in an entirely different manner than that in the closed case. One can see from Fig. 4 that at morning the humidities of the inside air is practically high to the ambient. As the inside air is heated in the noon to the level of the ambient, its humidities remain essentially the same for a few hours. The results show highest humidities in the kitchen and lower in the master bed room. Lower limit of humidity is around 80%. The range is slightly more for conditions with no or slow air movement and decreases with higher air speeds.

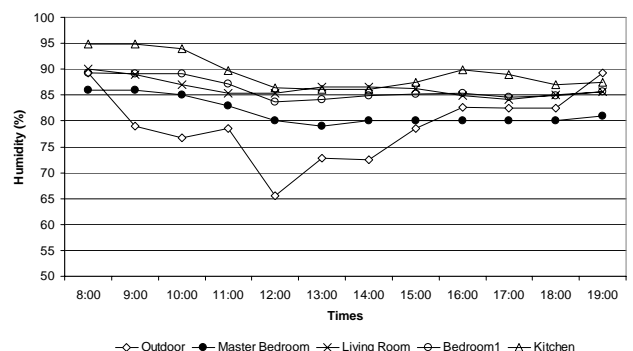


Figure 4. The Humidity Result of Measurements

3.2. Air Velocity, Mean Radiant Temperature and Predicted Mean Vote in Selected Room of Terrace house

Inspection of the thermal comfort performance results in master bedroom are presented in Figs. 5. First of all, the velocity measurements presented in Fig. 5 shows that a noticeable induced flow has been established inside the room. The velocity of this flow inside the master bedroom could reach the values of 0–0.001 m/s and even more. Comparison of the temperatures with the velocities in Fig. 5 shows no clear dependence of the air velocity on the temperature that small effect of the induced flow inside the room. One can see once again how the air temperature of the master bedroom, exceeds by about 0.5–1°C than temperature of the other room at noon. As in Fig. 5, here too the humidities were significantly higher than the ambient during the morning and afternoon hours. As a general trend, the mean radiant temperature (MRT) is lower than air temperature. In comparing both the MRT and the air temperature for similar comfort conditions at different air speeds the difference between air and MRT readings decreases with higher air speeds. As a variable against comfort votes (CV), MRT values show the best relationship as opposed to air temperature and relative humidity. The results of a 12 h run, presented in Fig. 7, show that during the afternoon hours the inside predicted meant vote (PMV) is noticeably lower (+0.8) than in the noon (+2).

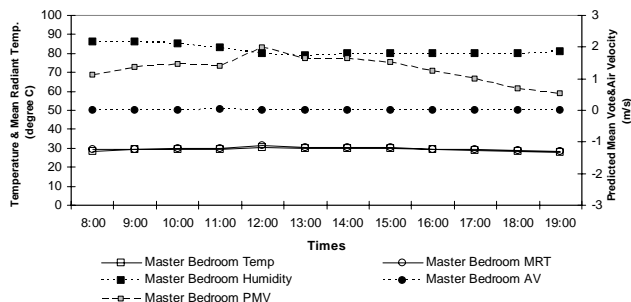


Figure 5. The Thermal Comfort Performances of Master Bedroom Measurements

The results presented above show that the decreased air temperature and increased air velocity provides better PMV inside the single storey terrace house. The air velocity and temperature distribution inside the master bed room are further illustrated in Figs. 6 and 7, where the numerical results are presented. Validation of the CFD model with the field measurement results was carried out in two steps. The first step was the validation of temperatures at various times and the second one was the validation of the air velocity. The purpose of this two-step validation was to find the cause of deviation between the two results

Fig. 6 showed a good agreement between CFD simulations with the field measurement results. The difference was less than 5% for most of the calculated points on the temperature and the relative errors were less than 2% except for the ambient temperature no differences. In those cases, the absolute differences were less than 0.4°C. In summary, the outdoor/ambient and indoor air

temperature calculated by the CFD simulation showed a good agreement with the results of the field measurement.

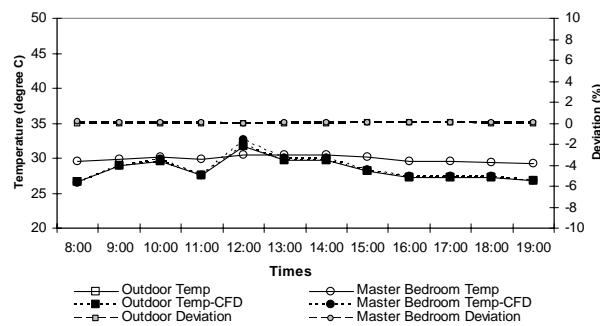


Figure 6. Comparison of the temperature result between field measurement and simulation

Fig. 7 shows the plot of points for field measurement and CFD simulation airflow velocity under outdoor and indoor conditions. The figure shows close agreement between measurement and simulation values. Deviations are within a range of 10% of the calculated air velocity. For most of the observations, this difference is less than 5% of the calculated values.

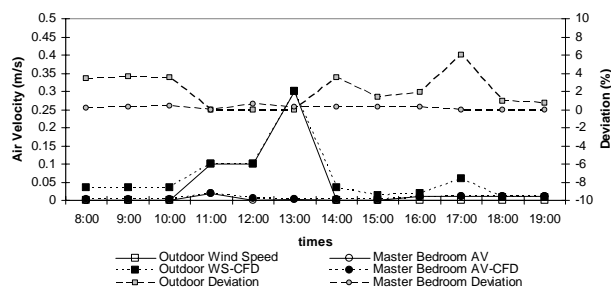


Figure 7. Comparison of the air flow velocity result between field measurement and simulation

Although there are slight differences between the results of CFD and field measurement, it can be inferred that the CFD modeling is appropriate to reproduce the phenomena occurring in the measurements. The use of the CFD model to investigate the performance of the thermal comfort ventilation can be validated.

3.3. Thermal Comfort Ventilation Condition

The predicted mean vote (PMV) index, proposed by Fanger, has been validated for predicting comfort in moderate thermal environment and used by ISO standard 7730 [6]. In this study, we used a new index predicted mean vote (PMV), as did in previous studies [1]. The new index (PMV) is proposed for any dry or humid environment by simply replacing operative temperature (T_o) in Fanger’s comfort equation with standard effective temperature [7]. Field studies conducted in warm climate has shown that the Fanger’s thermal model was not adequate for predicting thermal comfort under hot and humid climate in naturally

ventilated space or in building without air conditioning. It overestimates the thermal sensation of subjects [1]. In addition, based on physiological principles, people who have been living in warm environment prefer to feel warmer than neutral. Warm condition sensed centrally in the brain reacts stronger than cold receptor exists in the skin [8].

The subjective sensation of warmth or thermal comfort of the subject has traditionally been measured using a seven point scale where the comfort range is defined between -0.5 and +0.5 by the subjects' comfort votes [9]. A complete description of the new index PMV is described numerically as: cold uncomfortable (-2.5), cool unpleasant (-1.5), slightly cool acceptable (-0.5), neutral pleasant comfortable (0), slightly warm acceptable (+0.5), warm unpleasant (+1.5) and hot uncomfortable (+2.5).

To evaluate the PMV value in each room of the single storey terrace house, several simulations were performed (08.00 h–19.00h). In Fig. 8, a comparison between PMV measurement and PMV using CFD room is made. Comparison of calculated PMV using CFD and PMV measurement shows that the CFD thermal model is under estimate, especially in the afternoon. The result shows that the deviation of PMV results between CFD and measurement is about 0-20%. However, PMV measurement needs a higher air velocity to achieve comfort under activity of 1 met. PMV CFD for a master bed room indicated higher value than another rooms. It was found that at 12 am, the PMV is higher than other time. This condition is warm unpleasant. Therefore, decreasing PMV value is recommended to maintain the comfort level.

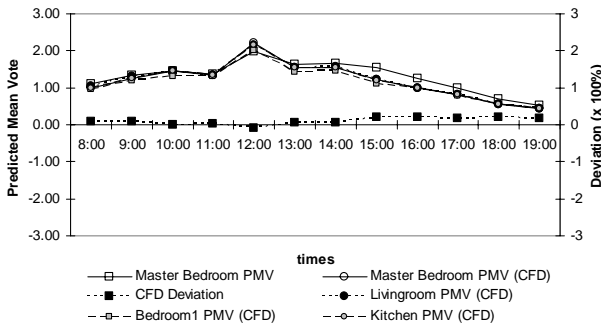


Figure 8. PMV Calculation of Single Storey Terrace House

A comprehensive study in trying to establish the Malaysian comfort zone was done by Mohd Rajeh Mohd Saleh (10). He used Auliciems' neutrality temperature equation (combining the free running and actively controlled buildings) to define the comfort temperature of Malaysians and proposed a bioclimatic chart showing the comfort zone for Malaysia base on Olygay's (1958), Arens (1984) and Szokolay (1997) version of the bioclimatic chart. With the width of the comfort zone taken to be 5°C (11) thermal comfort temperature extends approximately about 2.5°C above and below the neutral temperature. The upper limit of the comfort zone can be extended by providing natural ventilation or some form of air movement. With air movement of 1 m/s, the upper limit of Malaysia's comfort temperature can be extended by about

4°C to 33.2°C. Figure 9 show the proposed minimum air velocity for thermal comfort zone for Malaysia base on a bioclimatic version of the psychometrics chart (12).

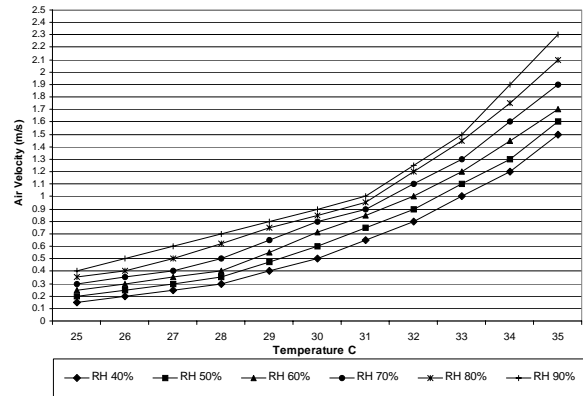


Figure 9. Minimum Air Velocity Requirements for Thermal Comfort Ventilation

Average indoor air velocities while seated (1.1 m), derived from simulation test, were found below the requirement of minimal air velocity (0.6 m/s-1m/s) for thermal comfort ventilation condition (temperatures: 26°C-33°C and humidities: 80%-90%). This suggests that the indoor velocities can increase to provide thermal comfort ventilation by modification in ventilation design.

3.4. Using Vertical Solar Chimney to Improve Indoor Thermal Comfort

The ventilation modification was introduced by the use vertical solar chimney ventilation (Fig. 10). It is found that in a terrace house, using vertical solar chimney can induce air to obtain natural ventilation. Results indicated significant increase in maximum air velocity (0.3 m/s) as well as cooling of temperature. The average temperature in the room ranges between 27.5°C and 31.1°C during 08.00 h – 19.00 h, which is about 0.5°C lower than a normal terrace house without solar chimney (fig. 11).

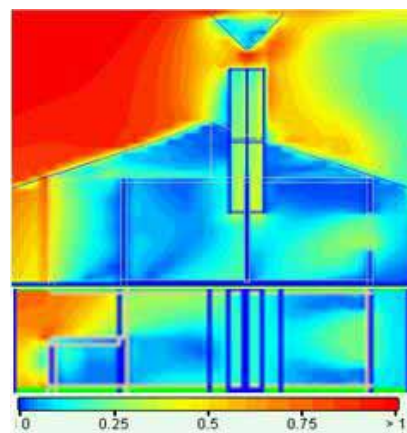


Figure 10. The Air Velocity Pattern in Vertical Section and Horizontal Plan (1.1m above floor)

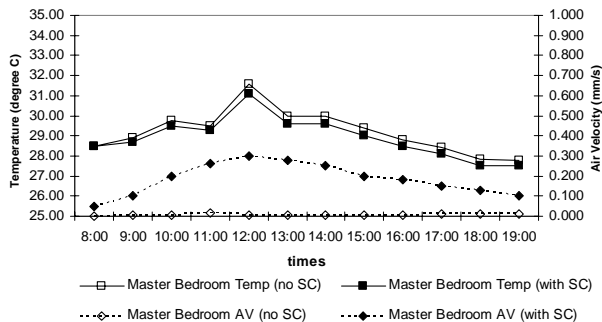


Figure 11. The Temperatures and Air Velocity Performances of Terrace House with Vertical Solar Chimney

4. Conclusion

The main goal of the study is to evaluate the thermal comfort ventilation performance of single storey terrace house. The tree evaluations were done: first, to compare between field measurement and CFD simulation, thus the simulation result showed good agreement with field measurements. The estimation of PMV thermal model by CFD simulation is lower than PMV measurement. It is proved by a higher airspeed needed to obtain comfort based on CFD thermal model. Second, to simulate the field study on selected day (in 21 March, at 08.00 h -19.00 h), that the air velocity result is inadequate to provide thermal comfort especially at 12.00 h. This study proved that the single side ventilation (in the case of master bed room) decrease the indoor comfort level. The last evaluation shows that natural ventilation to perform well under the ventilation modification by using the vertical solar chimney. The following issues should be considered in the design stage to obtain comfort in tropical humid climate:

- Low indoor air velocity and high ambient temperature may create unwanted negative ventilation by limiting the room design (single side ventilation) in terrace house.
- The use of vertical solar chimney ventilation to increase air velocity and improve the indoor comfort level under seating activity (1 met).
- The performances of ventilation design proposed in this study, if incorporated in terrace house model, is expected to create a reasonable indoor air velocity for human comfort, besides contribution to energy efficient and environmentally friendly options.

5. Acknowledgments Conclusion

The authors would like to thanks for Dilsan Remaz Ossen, Halimah Yahya, Lalu Mulyadi and Chia Sok Ling for help and support.

References

- [1] E. Prianto, S. Houpert, P. Depecker, J.-P. Peneau, , 2001, Contribution of numerical simulation with SOLENE to find out the traditional architecture type of cayenne, Guyana, France, International Journal on Architecture Science, Hong Kong 1 (4) 156–180.
- [2] P.O. Fanger, 1972, Thermal Comfort, McGraw-Hill, New York, 244 pp.
- [3] H.B. De Wall, 1993, New recommendations for building in tropical climate, Building and Environment 28 (3) 271–285.
- [4] B. Griefahn, C. Ku`nemund, U. Gehring, P. Mehnert, 2000, Drafts is cold environments. The significance of air temperature and direction, Industrial Health 38, 30–40.
- [5] Ansley, R.M., Melbourne, W and Vickery, BJ. 1999. Architectural Aerodynamics. London : Applied Science Publishers Ltd.
- [6] ASHRAE Handbook Fundamentals, 1989. American Society of Heating and Air Conditioning Engineers, Inc. Atlanta.
- [7] A.P. Gagge, A.P. Fobelet, L.G. Berglund, 1986, A standard predictive index of human response to the thermal environment, ASHRAE Transactions 1986 (Part 2B) 709–730.
- [8] E. Mayer, 1993, Objective criteria for thermal comfort, Building and environment 28 (4) 399–403.
- [9] M.A. Humphreys, J.F. Nicol, 2001, The validity of ISO-PMV for predicting comfort votes in every-day thermal environments, Moving Thermal Comfort Standards into the 21st Century, Windsor, UK, Loughborough University, pp. 406–430.
- [10] M Rajeh. 1989. Natural Ventilation in Terrace Housing of Malaysia : Effect of Air Well on Air Flow and Air Velocity, University of Queensland, Master Thesis.
- [11] Szokolay 1997, Thermal Comfort in the Warm-Humid Tropics, ANZASCA Conference, Brisbane,
- [12] Nugroho, Agung Murti and Hamdan, Mohd. 2006. Evaluation of Parametrics for The Development of Vertical Solar Chimney Ventilation in Hot And Humid Climate, The 2nd International Network For Tropical Architecture Conference, At Christian Wacana University, Jogjakarta,

The Development of Mosque Architecture in Malaysia: Classification on Possible Styles and Influence

Mohamad Tajuddin Mohamad Rasdi¹, Nangkula Utaberta^{2*}

¹ Prof. Dr. Department of Architecture, Faculty of Built Environment
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5536544, E-mail: mindakritis@yahoo.com

² PhD Candidate, Department of Architecture
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-7-5530511, E-mail: nangkula_arch@yahoo.com

Abstract

The main purpose of the paper is to set up an initial framework of architectural description of Muslim architecture in Malaysia within the context of the architectural concern of 'style'. The main concern is to elucidate clearly the different architectural styles of Muslim architecture and suggest hypotheses for future studies in relation to the forces which might have been instrumental in the conception of these styles. The paper is important in order to project the message that much of what we understand as 'Islamic Architecture' in the present literature comes from various artistic and political agendas that have questionable merits compared to an analytical approach of Islamic architecture from the Qur'an and the Sunnah of the Prophet Muhammad (peace be upon him). The paper is also important for practicing architects to understand that the present syntax of 'Islamic Architecture' with its references in Middle Eastern, Mid Asian and African models are not necessarily the best precedence to project the architectural message of Islam. As opposed to the three classification of Traditional, Colonial and Modern styles of mosque architecture, the paper has identified seven main styles with a suggestion of the rationale and influences of those styles.

Keywords: Mosque, Style, Islamic-Architecture

1. Introduction

1.1 The Idea of Islamic Architecture

There are many questions that are related to the idea of what constitute 'Islamic Architecture'. The most prevailing view seems to be that Islamic architecture possess a set of architectural vocabulary that is indisputable. The most consistent perception of Islamic architectural vocabulary is represented by the many monumental state and community mosque that uses the Middle Eastern or Central Asian architectural garb as dutifully exemplified in its lavish splendor as in the Wilayah Mosque in Kuala Lumpur and Putra Mosque in Putrajaya. The present paper seeks to establish that this idea of a consistent architectural vocabulary is not what Islamic architecture is all about. We wish to show that Islamic Architecture has embraced many forms and vocabularies throughout the centuries in response to climatic, technological and socio-political concerns. The focus of this study is in Malaysia and we have identified eight different major styles of architecture each with its probable socio-political rationales. This study is important to pave the way of a more enlighten intellectual discourse towards a progressive theoretical construction of the idea of 'Islamic Architecture'.

1.2 Paper Scenario on Mosque Architecture

There is not an abundant literature on the subject of

mosque Architecture in Malaysia. However there are five types of writings and concerns surrounding the meaning and description of mosque architecture in this country.

The first type of writing is more concerned with historical description of mosques in Malaysia. Foremost in this area is the writings of the late Abdul Halim Nasir. He was an avid traveler and a prolific writer attached to the National Museum of Malaysia. He has traveled widely in the Malay World and have taken many photographs and provided quite a number of sketches, oral traditions and his own observations of mosques in this region. There are generally two problems in his writings. Firstly, being trained not in the field of architecture, much of his analysis, description and classification particularly that of the modern mosques are wanting in the aspect of modern architectural theory. Secondly, his academic rigors in presenting comparative analysis of typologies and description is limited to a learned persons first hand account and not a systematically and scholarly exercise. We wish not to belittle his work because his contribution to the foundation of Malay architecture is monumental and indisputable.

The second type of study is a more rigorous form of architectural description as found in our writing 'The Architectural Heritage of The Malay World: The Traditional Mosques'. This book contains measured drawings and detailed oral traditions of six traditional mosques with a comparison of their architectural styles. All of the empirical

data was taken from the Center of Built Environment in the Malay World or KALAM (*Pusat Kajian Alam Bina Melayu*) based in the Faculty of Built Environment, Universiti Teknologi Malaysia. However this book is lacking in the examination of the many the historical sources in the Malay World.

The third type of study is related more to a cursory historical description and a tinge of modern theory construction as found in the works of the architect David Mizan Hisham. As a practicing architect, Hisham writes well in relation to a typological approach in his description but lacks sorely in the historical and religious backing of many of his statements. The problem of typological writings is that it schemes the surface of architectural design and merely looks at the outward form as a conclusion without knowing its content.

The fourth type of writing is concerned primarily with determining the iconography of mosques in Islam and thus attempts to establish a theoretical design idea for that building. One author that is famous for his esoteric suggestions of the meaning of mosques come from Nakula. We feel that he relies too much on Sufistic ideas at the expense of other interpretation of Islam. His works suffers much the same problems of Syed Hussein Nasir and Titus Butckhardt simply because he seems to rely much of his foundation paper on these two scholars.

The fifth genre of writing comes from our work in attempting to establish a design formula for mosques by using the socio-political analysis of the religious sources of the hadiths and the Qur'an. In our books, 'The Mosque As A Community Development Center' and the book '*Peranan, Kurikulum dan Reka Bentuk Masjid Sebagai Pusat Pembangunan Masyarakat*' (Role, Curriculum and Design of Mosques as a Community Development Center) we have shown how Western architectural and historiographical perspectives have wreaked havoc on our understanding of the mosque in relation to its role and architecture. We have also hinted at the problem of Sufistic and Shia'istic framework in many writings of mosque architecture that have blindly been accepted by clients and architects in their justification of any design ideas. In our work we have relied solely on the religious sources of the hadiths and the Qur'an interpreted within the sunni perspective of Islam and have derived the eternal idea of the mosque in Islam. This idea is then combined with the considerations of the contemporary societal problems of Muslim communities to produce an activity and an architectural blueprint of mosque development. The present paper attempts to fill in many of the gaps left in the five writing genres as stated above. What we seek is merely a more detailed classification and a suggestion of probable reasons that might have led to the decision to design mosques as such in its own historical period.

2. Mosque Architecture in Malaysia: Classification of Styles and Possible Influence

3.1 Introduction

There are seven styles of mosque architectural language that can be classified. They are the Traditional Vernacular,

the Sino-Eclectic, the Colonial, the North Indian, the Modern Vernacular, the Modernistic Expressionism, the Post-Modern Revivalism. This section describes the range of characteristics of each style together with the rationale for the choice of names used. The section also contains suggestions of socio-political influences that might have made significant impression on the choice of styles used.

3.2 Traditional Vernacular Style

The word traditional used in this context represents practices and ideas of Malays before the colonialist came to the country. The word vernacular denotes the availability of materials, craftsmanship and technology of the pre-colonial period.

There are three types of mosques in this category which can be distinguished mainly from the use of the roof form. The first is the three tier pyramidal roof form as in that of Kampung Laut and Kampung Tuan believed to be built between the sixteenth and the seventeenth century. From the tip of the pyramid to the bottom of the columns, the mosques can be inscribed into an almost perfect cube. The second type is the two tier pyramidal roof forms such as that of the Papan and Lengeng Mosques. They are built between the eighteenth and nineteenth century. The third kind is the gable roof form which is similar to that of the house. There are the single or double tier roof form in this typology. All of the mosques are raised from waist to shoulder height.

All of the mosques in this stylistic classification are made of timber. All of the mosques originally possess no serambi area. Many of the serambi or verandah came later. An interesting feature of these mosques is that they do not come equipped with a minaret. All are added later whilst some mosques are still without any minaret. All of the mosques display only a single space layout and no indication of subdivision can be found. Since many of the mosques are built close to a river for ablution purposes few of them have wells for preparation of prayer. There is no indication that these early mosques were fenced up and isolated from the villages.

All of the mosques use the traditional timber post and beam system. Columns on the exterior and the interior are common features except in the Kampung Tuan Mosque. In this particular mosque, there are only exterior column without the traditional four or five internal pillars. A close examination reveal that this particular mosque uses a unique timber 'space frame structure but known locally as the tiang gantung system. It is a fact that all the traditional Malay construction do not use any form of truss system except the occasional king post which is denoted as the 'tiang gantung'.

The early patrons of mosques were religious scholars and there may have more say in what the mosque should look like. Since many of them had one time or another studied in many centers in the Asian region, they must have brought along their architectural influence. Although there is another hypothesis that the form is a result of Sufi mysticism of the ascendancy of the spiritual passage, there is no strong evidence except the simple conjecture of anthropometric association or physical ascension.



Plate 1: (left to right) Masjid Kmapung Laut, Masjid Kampung Tuan, Masjid Lenggang, Masjid Langgar

3.3 Sino-Eclectic Style

The term Sino Eclectic is derived from two aspects of the phrase. The term Sino indicates a Chinese influence and eclectic which denotes a combination of two or more influence of architectural language. There are two types of mosques in this category. The first is the three tiered pyramidal roof form and the second is the double tier pyramidal roof form. Both types are similar in a majority of other features. The first type is characterised firstly by the three tiered pyramidal roof form similar in proportion to that of the traditional vernacular style. The differences between the two styles lie in the prominent curvature of the roof ridges which is made of cement. The mosques of this style are such as the Kampung Hulu Mosque, the Tengkeri Mosque and the Kampung Keling Mosque which are all in Melaka. There is also the Undang Kamat Mosque and a few others similar to it scattered in Negeri Sembilan. The Lebuah Acheh Mosque in Penang is also of this style. The mosque of this style sits on the ground and are not raised like the traditional vernacular ones. They all have slabs on grade which are raised about half a metre high with stone stairways accessing to the main floor plan. The plan of the mosque proper consists of the enclosed prayer area and the *serambi* or verandah surrounding either three parts of the square plan or all around it. These mosques are all located in the dense urban areas and the building is surrounded by a masonry fence with sometimes a roofed gateway like almost reminiscent of Chinese temples. The ablution area is taken care of by a masonry pool and there is almost always a cemetery patch as part of the mosque compound. In later times madrasah or religious schools and imam's residences are built flanking the mosque proper.

As opposed to the *ulama* sponsored patronage of the traditional vernacular style, the sino eclectic mosques have Chinese Muslim Merchants as their patrons. The rise of the merchant class saw the country in a situation where these wealthy people wield considerable influence over the socio-political affairs of their community. Since many of the mosques were built by Chinese craftsmen, there is the possibility of suggesting the similarity of Chinese architecture. The gateways are also elements that reinforce the notion of architectural language transfer from the Chinese religious architecture.



Plate 2: (Left to Right) Masjid Kampung Hulu, Masjid Kampung Keling, Masjid Tengkeri, Masjid Dato' Undang Kamat.

3.4 European Classical Style

European Classicism refers to the High Renaissance architecture that was derived from the Greco-Roman heritage. The main characteristic features are the use of the

definitive tripartite division of base middle and top with double column supporting semi circular arches or walls with pilasters. A strong symmetrical composition of the massing and space is also an identifying feature. The Sultan Abu Bakar Mosque in Johor Bahru is equipped with four stout minarets with small domes crowning the top. The main prayer hall has a closed hipped roof. The European Classical style is distinguished by its elaborate exterior cornice work forming a continuous band around the building. The windows are also framed by plastered cornice work. The Pasir Pelangi Mosque uses a deep pyramidal roof form reminiscent of the traditional vernacular mosques of the past but without any dividing tiers. The minaret is heavy in proportion and capped not by a dome but by a small pyramidal roof. The Sultan Ibrahim Mosque in Muar, Johor uses a hipped gable roof which covers a large central portion of the prayer space. This part of the mosque protrudes above the rest of the roofed area to form clerestory windows reminiscent of the basilican type churches of early Christianity in Rome.

It is most curious to note that all the mosque sampled are from Johor. One hypothesis might be that the colonial masters wished to present an idea of a more sedate and rational Islam as opposed to the radical voices of fundamentalist Islam calling for the ouster of the colonial rule. However, a stronger hypothesis might be derived from the simple fact that the Sultanate of Johor shows no secret for their desire to be culturally Western from the English Cottage architecture chosen for the royal palace and residence of Johor at Bukit Sirin. A more glaring clue is the other palace fronting the Lido beach in downtown Johor Bahru which is dressed purely in 18th century Classical garb. There may perhaps be the idea of 'westernizing' Islam to be more like the English understanding of Christianity.



Plate 3: (left to Right) Masjid Abu Bakar dan Masjid Muar.

3.5 North Indian Style

The term is used to describe the imitative Moghul type architecture that once flourished colonial Malaya. The North Indian style is easily distinguished from any others due to its generous use of small and large onion domes, multitude of spires and small domed canopies, more than one minaret, horse shoe or multifoil arches over decorated columns. The Ubudiah Mosque in Perak, the Kapitan Keling Mosque in Penang, the Jame' Mosque and Masjid India in Kuala Lumpur are prime examples of this style. The plans of the mosques are similar to the mosques of any other style with its central domes and arched verandah way. The mosque compound is fenced up at the full perimeter with the usual parking and grass area.

The early twentieth century mosques were built of masonry structure and enclosure system. The roofs are covered with masonry domes and others are of corrugated asbestos sheets. There are two possible reasons for the

proliferation of this style. The first is the colonialist choice of 'Islamic Architecture' to perhaps pacify the people into accepting their version of ritualistic Islam. This is evident in the government buildings such as the courthouse and railway stations having the same architectural language. Since the British had a strong hold over India, it seems a natural choice for them to help define Islamic Architecture from that region of the world. There is also another possibility that England during the 19th century was also bitten by the Picturesque bug. The residence of the great English architect, John Nash was built of the same North Indian Moghul style. This exotic preference might have been an obvious choice to frame the idea of a non-rational or exotic religion as opposed to the strict rationality of the Classical language reserved for the Churches and public buildings in England. The second probable reason for the proliferation of this style is the rise of the Indian Muslim merchant class. In Taiping and Penang, for instance, much of the city's commercial power was not also held by the Chinese but by the Indian Muslim merchant community. As these merchants prospered, they went back to their motherland and that might explain why the North Indian style is no longer favored for mosques since the rise of the Malay middle class saw no reason to associate Islam with native India.



Plate 4: (Left to Right) Masjid Kapitan Keling, Masjid India, Masjid Jamek, Masjid Ubudiah

3.6 Modern Vernacular Style

The term modern vernacular is used to refer to buildings which are constructed using mainly reinforced concrete structural frame with plastered brick infill. It is considered 'vernacular' because the practice and availability of these new materials and construction technique is the norm in this country presently. With respect to the building form and architectural language, mosques in this category would usually portray a gable or pyramidal roof with a small dome or the more expensive ones would find a larger single dome over the main prayer space. These types of mosques are the ones used in most of the modern housing estates in this country. Mosque within a 'modern village' or a squatter village with timber or semi timber houses would usually have a simple gable roofed mosque of timber or partly built of masonry. The verandah would be enclosed by a series of arches. The plans of these mosques usually contain a single prayer space with the women area being screened off on the rear part of the space with a curtain or movable partitions. It is obvious to notice that the women space seems to be an after thought in most of these housing estate mosques. Most of these mosque would have one or two minarets with sometimes a grand portal gateway where the communities there are well off or have close political connections with the ruling party.

The Haji Abdul Hadi or Rusila Mosque in Terengganu and The Nik Aziz Mosque in Kelantan differs in that they are

used as madrasah or religious schools where the two Chief Ministers teach the people daily. These mosques possess an aura closer to that of the Prophet Muhammad's early mosque as they are used as community centers and places for travelers to rest overnight. The facilities include libraries, shops, student dormitories, travelers' lodgings and administration center. It is also important to note that these two mosques do not possess a definite fence but the grounds and that of the surrounding houses seemed to be shared.



Plate 5: (left to right) Masjid Rusila, Masjid Kampung Batu, Masjid Sentul.

3.7 Modernistic Style

The term 'modernistic' is derived from the ideas and main principles that govern the early revolution of architecture in Europe in the early twentieth century. Much of these ideas relate to notion that a true architecture is one that rejects historic revivalism in any form, that rejects ornamentation in any form, that celebrates abstraction in forms and that celebrate the structural expression in architecture. There are two types of modernistic styles in mosques of Malaysia; the Modernistic Expressionism and Modernistic Structuralism. The phrase 'modernistic expressionism' is derived from William J. Curtis' classification of 'expressionism' as any form of architecture that carries a metaphoric message through the use of structurally expressive form. Eric Mendhelson's Einstein Tower as well as Eero Saarinen's TWA Airport Terminal are examples of these kind of architecture. I have added the word modernistic because in some buildings only one part has the expressive qualities whilst the other parts are those that subscribe to the general modern architectural language. There are two mosques in this category in Malaysia. The Masjid Negara and the Negeri Sembilan State Mosque are the only two examples of this style. The Masjid Negara is the best example of the combination of a modernistic reinterpretation of traditional Malay Architecture with a folded plate 'dome' with a metaphor of a royal umbrella signifying the importance of the building as a national monument. The Masjid Negara uses an extensive set of serambi or verandah space with light courts and air wells to provide ample daylighting and passive cooling to the building. It is by far the best example of a building imbued with the technological and spiritual qualities of an architecture with a true Malaysian identity. The Negeri Sembilan State mosque uses a series of intersecting reinforced concrete conoid to refer to the horn-like gable roofs of the Minang traditional architecture. The reference to the 'bubungan gonjong' is uniquely expressed in the structural play of the conoids. The architect had not resorted to the simplistic revivalism alternative of the traditional roof but has reinterpreted it in an abstract but creative way.

The other type of style within this category is the Modernistic Structuralism Style. This is the classic Miesien tradition of treating the building as a mere machine of structural expression, nothing more, nothing less. The

dictum of 'less is more' echoes through the buildings of this style. The Penang State Mosque presents a concentric ring of curved reinforced concrete ribs. The tip of the rib is crowned with an awkward dome to give its 'Islamic' signature. The Kota Samarahan Mosque in Sarawak has an identical form with the exception that it uses steel delta trusses with stretched teflon tensile fabric as the roofing material. The Al-Syahidin Mosque in Sik uses a folded space frame structural system anchored at four points to the ground. The roof spans a space totally and uniquely devoid of any solid wall. The Qibla wall is a free standing structure whilst the whole floor is ringed with a meter high railing. These mosques speak no particular meaning other than a simple expression of basic shelter. As indicated the main structural system used reinforced concrete frames and shell. The folded plate and intersecting conoidal shells are seen as progressive images that can be attached to the dynamism of Islam. There is also a number of buildings of steel as the main structure with metal decks or Teflon stretched tents as the roofing. The form of the structures can be said to be borne solely by the architects because clients at whichever level of society are unable to understand or converse in the structural language and intricacies of structural metaphors.



Plate 6: Masjid Negara, Masjid Negeri Sembilan, Masjid Al-Syahidin, Masjid Samarahan.

3.8 Post Modern Revivalism

The term post modern denotes an approach that contradicts the principles and edicts of what was understood as the modern style. The term revivalism denotes one of the many ways which the post modernist attempts to create an architecture of meaning for the general public rather than for the elite few. There are two kinds of Post Modernist Revivalism in this category of style. The first is the Foreign Revivalism and the second is the Vernacular Revivalism. Foreign Revivalism in mosque architecture seems to be the order of the day with such examples as the Putra Mosque, the Shah Alam Mosque, the Wilayah Mosque, the UTM Mosque in Johor, the Sarawak State Mosque and many others. These grandiose statements of so called 'Islamic Glory' is the preferred language by state and federal governments to express their 'deep concern' over the important symbol of Islam. The use of an eclectic array of Iranian and Turkish domes, Egyptian and Turkish minarets, Persian Iwan gateways, lavish courtyards surrounded by the sahn, and an Arabian hypostyle planning composition and pointed or semi circular arches bathed in sumptuous classical 'Islamic' decorations are the vocabulary of mosque design. The Vernacular Revivalism calls for a slightly less monumental approach with its use of the three tiered pyramidal roof form built either of concrete or timber. The Melaka State Mosque represents the grandiose extreme whilst Jimmy Lim's ASPA mosque for the villages of Pahang represents the other more humbler version. The Melaka State Mosque combines the

use of arches and gateways along with its neo-vernacular imagery whilst the ASPA mosque caters more strictly to the scale of modern timber construction without the flavour of any Middle Eastern or Central Asian touches. With the exception of the ASPA mosque, the others are fenced up complexes with lavish compounds filled with fountains, paved grounds and grass lawns with much sculptural landscaping. The siting of these mosques are on top of the highest points, in the middle of man made lakes or isolated from the urban fabric.

The mosques are mainly of reinforced concrete with steel ribbed domes. Marble and precious imported tiles envelope the whole mosque. The ASPA mosque is refreshing in its use of modern timber trusses and louvered walls with a perimeter railing guarding the surrounding *serambi* space. It does not have any fences and blends humbly with its surrounding village atmosphere. An interesting question about the Foreign Revivalism style concerns the departure from the British lead Moghul venture. It may be the desire of the politicians and architecture elite to depart from any reminders of British colonialism. It may also be the idea of Pan Islamic resurgence as in the rise of the Iranian Revolution and various revivalist movements in Egypt. It would certainly be easier for the ruling party to show its appreciation for the revivalist fundamentalist movement of Islamic resurgence in the form of masonry, concrete and steel rather than its actual realization of these movements in respect of moral values and political models. The construction of the Melaka State Mosque is actually a response to the historical site of the state where the Melaka Mosque model is obvious. However, the construction of the ASPA mosque is one of the first inklings of a criticism against the status quo mosques in preference of a regional identity of Islam.



Plate 7: Masjid Shah Alam, Masjid Wilayah, Masjid Putrajaya, Masjid ASPA, Masjid Melaka.

3. Conclusion

There is a varied architectural style for mosque design in Malaysia. One glaring conclusion is that the idea of a definitive 'Islamic Architecture' vocabulary leaves much to be desired since there does not seem to be a preference for any one particular style that is repeated. The advocates of Middle Eastern revivalism must admit that though Islam may have been given birth there, that fact is in no way a strong support for the style's monopoly over Islam. If anything, one can conclude that Islam allows the variety in architectural language because of its principle religious

tenets that is beyond racism or parochialism. The varied style is a testament to the adaptability of Islam to the various cultures and belief system that does not contradict its main focus of worshipping one God. Another conclusion that can be drawn is that almost all of the style are in one way or another imitative of outside influences where Islam had flourished and had become a strong political force. For instance, the use of the pyramidal roof form may be construed as the imitation of Javanese mosques where Islam had become a strong political force. The use of the North Indian style again testifies to the revolutionary spirits of Islam in that region that ultimately led it to the setting up of the Islamic republic of Pakistan. The Iranian and Egyptian architectural preference in contemporary state mosques are symbols of the success of Islam as an awakening force in the two countries. The Modernistic Structuralist style does not seem to possess whatever political agenda other than a mere whimsical play of new toys as in the space frames and tent structures. However the wall less Al-Syahidin Mosque is a brave attempt of reinterpreting the idea of the mosque in its basic and pristine form...that of a temporary shelter in this world. The final conclusion that can be drawn from the survey and classification is the undoubted fact that the Masjid Negara holds a unique and inspiring position as the Mosque that does not have any inferiority complex about which place and culture does Islam seem to be best represented. Its honest effort at interpreting a progressive identity whilst answering the call for a building of humble cost and not of astronomical extravagance, the mosque displays a tropical composition to be proud of. It would be fitting to end on the note that architects and politicians should be aware of the remarkable achievement of this national monument in order to properly present Islam in its true form without the baggage of intellectual regressivism, inferiority complex and parochialism.

Reference

1. Abdul Halim Nasir. 1995. *Seni Bina Masjid di Dunia Melayu-Nusantara*. Bangi: Universiti Kebangsaan Malaysia.
2. Ali, K. 1976. *The Vanishing Heritage*. Tesis Sarjana, University of York.
3. Blake, P. 1976. *The Master Builders*. Alfred A. Knoff, New York.
4. Curtis, W. J. 1982. *Modern Architecture Since 1900*. New York: Phaidon Press Incorporated.
5. Cresswell, K.A.C.. *A Short Account Of Early Muslim Architecture*. Beirut: Librarie du Liban, 1968.
6. Collin, Peter. *Changing Ideals In Modern Architecture*. London: Faber and Faber, 1965.
7. Gibb, H.A.R. (ed.), *The First Encyclopedia of Islam 1913-1936* E.J. Brill, 1987.
8. Guillaume, A. *The Life Of Muhammad: A Translation Of Ibn Ishaq Sirat Rasul Allah*. Oxford: Oxford University Press, 1982.
9. Hoag, J.D. *Islamic Architecture*. London: Faber and Faber, 1989.
10. Jencks, C. 1977. *The Language of Post-Modern Architecture*. London: Academy Editions.
11. Kleinbauer, Eugene W. *Modern Perspectives Of Art History* New York: Holt, Reinhart & Winston Inc. 1971, pp. 37-105.
12. Masjid Kampung Laut, Kota Bharu, Kelantan' Unpublished Report from the KALAM archives.
13. Masjid Langgar, Kota Bharu, Kelantan' Unpublished Report from the KALAM archives.
14. Masjid Kampung Keling, Jalan Tukang Emas, Bandar Melaka,' Melaka Unpublished Report from the KALAM archives.
15. Masjid Kampung Hulu, Bandar Melaka,' Melaka Unpublished Report from the KALAM archives.
16. Masjid Kampung Tuan, Kemaman, Terengganu' Unpublished Report from the KALAM archives.
17. Masjid Undang Kamat, Tanjong Ipoh, Negeri Sembilan' Unpublished Report from the KALAM archives.
18. Mawdudi, Abul A'la. *Let Us Be Muslims* Leceister: The Islamic Foundation, 1992.
19. Mawdudi, Abul A'la. *Towards Understanding Islam*. Translated and edited by Khurshid Ahmad. Revised edition. London: The Islamic Foundation, 1989.
20. Yeang, K. 1992. *Architecture of Malaysia*. Amsterdam: Pepin Press, 1992.
21. Yoong, C. C. (penyunting) (1987). *Post-Merdeka Architecture Malaysia 1957-1987*. Kuala Lumpur: Pertubuhan Arkitek Malaysia.

Redevelopment The Malay Agriculture Settlement (MAS) Areas of Kampong Baru Using Land Readjustment Technique

Prijono Nugroho Djojmartono¹, Djurdjani², Ismail Omar³, Asiah bte Othman⁴

^{1,2,3,4} Department of Land Administration and Development
Faculty of Geoinformation Science and Engineering
Universiti Teknologi Malaysia

Abstract

The phenomenon of underutilized built-up area located adjacent to tall modern city is appeared in many developed countries. The MAS areas of Kampong Baru, in the capital city of Kuala Lumpur, is one of the example. The redevelopment of this area is not only creating the urban land supply but also revitalizing the inner cities. Multiple ownership, less attractive investors, and land policy constraint are identified as the main problems in redeveloping this area. The land readjustment technique is considered as an alternative approach for solving this problems. This paper describes the concept of land readjustment technique and subsequently studies, and the possibility to implement it in Kampong Baru based on the previous experiences in many countries.

Key words: land readjustment, land supply, urban redevelopment

1. Introduction

Land is the base for all the development activities. The urban land is more valuable than rural land, because the concentration of activities is relatively higher than the rural areas. There is a competition between the activities for space and therefore the supply of land is relatively inelastic in urban areas. The demand for land is increasing at a very rapid rate in virtually every urban center throughout the world. According to this demand that arises day by day, the activities locate here and there, creating an inefficient urban system. This is known as the misuse of urban land (Cunningham, 2005). Congestion at low levels of density, urban sprawl, vacant lots, and underutilization are all manifestations of this phenomenon. This ultimately results on the restriction of the supply of land. Any constraint to the supply of land negatively impacts on the human settlement development (Nishani, 2003). But planning and management can play a significant role in the fulfillments of future land needs. Planning can also directly influence the demand for land. Indeed planning potentially has the ability to orchestrate a number of tools so that land supply and land demand can be brought into balance. Land readjustment is one of the techniques that has been tried in various countries and has obtained successful results.

The objectives of this paper are: to study the process of land readjustment; to examine the experiences of countries where the technique has been applied; and to discuss the potential application of land readjustment as a redevelopment technique for the MAS areas of Kampong Baru.

2. The Land Readjustment (LR)

The LR is one of techniques and procedures for urban development, mainly focuses on the modification of existing land parcels in terms of shape, location, size and other conditions together with the proper allocation and improvement of urban infrastructure (Hayashi, 2002). The LR is principally not adopted the purchase system for the land acquisition. It is a land title conversion by means of amalgamation, subdivision, relocation, exchange, and modification of size and shape of land parcel based on new layout plan and its physical improvement. The conversion of land title is made principally based on the land utility value increase, and the project cost is largely recovered by increased land utility value through the process of land contribution. The land contribution, in practice, is made within the total amount of land development profit, and shared among the total landowners in justice manner. As the name implies, the land contribution is made by land itself, but not by the cash payment. The landowners are nominally benefited through the project. Thus every landowner is requested to contribute a part of land for the recovery of the project cost as well as the land needed for the infrastructure sites (Hayashi, 2002). The LR can be applied to a variety of urban development including suburban new development as well as the redevelopment in the built-up area. The LR also can apply to a large-scale development such as new towns, and also to a small scale urban development less than 1 hectare as well. The LR could be applied to a variety of the urban development project, but the effectiveness and efficiency of the project differ to place to place affected by existing land use conditions and landowners perception.

2.1. Methods of LR Project

The mechanism of LR project is based on a series of land manipulation such as land contribution, replotting and land re-registration. The construction work and infrastructure improvement are done almost same procedures and manners of other urban development project. The specific aspect of LR project is land replotting that results in land contribution and creation of reserve land. The land is not purchased for the project, but all land titles are basically retained as previous conditions, and will be replotted in different location, shape and conditions according to the new layout plan and its development (Hayashi, 2002). The landowner involvement and participation in the LR project, the method of financial recovery and land replotting without the purchase are characteristic features of the LR project. Compare to ordinary private development, the LR has different land manipulation and financial systems for urban development. The heavy burdens of land purchase is avoided in the LR project, therefore, development area could be expanded in wider area in term of financial capability rather than normal land purchase system. The LR finance is principally accrued from the land contribution made by each landowner, and also to utilize the available funds as government subsidies, cost sharing or co-financing for the major infrastructure improvement (Hayashi, 2002).

2.2. Procedures in LR

Figure 1 provides an indication of a typical LR project. Before the project performed, the plots may have a different physical shape as well as a different value. Some plots may be hilly; some may be undulated and not suitable for constructions. Similarly some plots may be located closer to infrastructure hubs and therefore associated with a high value. However after the project the size of plots is reduced to obtain sufficient land for provisions of roads, infrastructure, and social facilities, open spaces and to reserve space for sale. According to the Figure 1 before LR the parcels were irregular in shape. When the plots are assembled the total extent of the area is 2455m². The area for public uses is then determined by measuring the square meters in the planned streets, parks, comparing it to the total area of the project (Doebele,1982). In other words, all land parcels within a project area are grouped together and a percentage of each land parcel calculated to determine the contribution to public areas. This percentage depends on the size of the project area and the total size of required public-use areas. The remaining land is then reallocated within the other blocks. To do this, first each site block is subdivided into suitable new lots, then land re-distribution is carried out. The basic principle in the distribution is to keep land in nearby its original location at least in the same block.

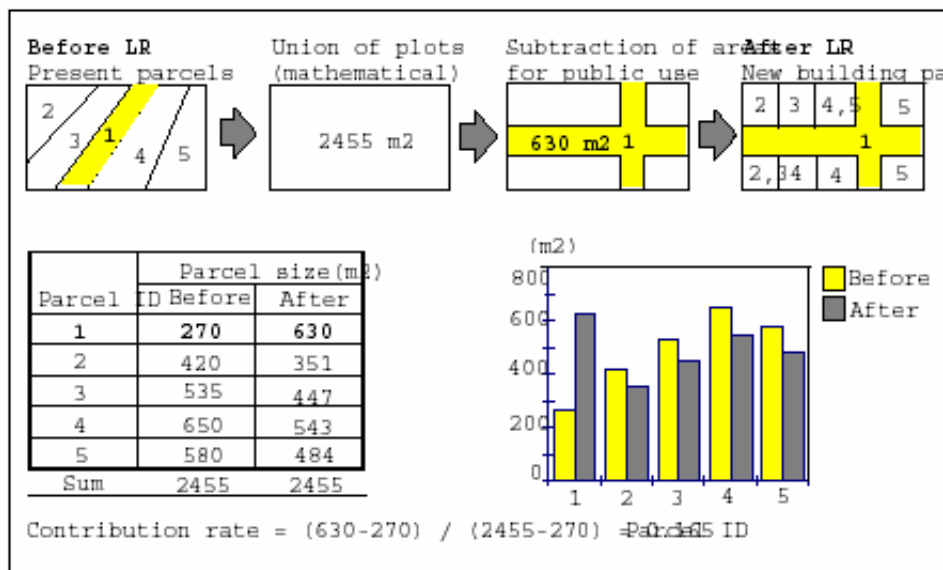


Figure 1 Procedures in LR (source: Doebele, 1982)

2.3. The benefits of LR

The benefits of LR for the government are:

- a) Compensation expenses for public-use land are greatly reduced and the public areas are captured in a more economical way.
- b) The ownership of land is clarified and through this the government revenue can be increased from property taxation.

- c) A zoning plan is realized in a short time, and it ensures a planned development soon.
- d) The existing cadastral records are updated and reorganized.

The benefits of LR for land owners are:

- a) Land value increases and land become more valuable
- b) The plot is transformed into a sufficient site that can use economically

- c) Fragmented small parcels are consolidated into buildable plots.
- d) Basic public services are received.
- e) No extra charges to land owners for the project expenses

2.4. The experiences of LR

In Germany, where modern methods of LR began in 1902, originating with the Lex Adicke in Frankfurt, the primary objective of LR was the assembly of small parcels when required for more efficient agriculture, for better urban development, or for other public purposes related carrying out an urban plan.

Experiences in city of Nagoya

Nagoya is one of the few planned cities in Japan. But the city is a pioneer in LR schemes. Nagoya city was seriously damaged during World War II as about a quarter of the city area was burnt down. The city rose to the present state through careful planning with public participation. Nagoya use LR techniques to carry out planned urbanization. Other cities have also benefited from these techniques since Nagoya has provided training programs for the engineers and planners (Viitanen, 2002).

Experiences in the Republic of Korea

LR programs, first introduced in Korea in 1934, have been actively pursued in the cities of Korea since 1937 as a viable means of modernizing the urban spatial structures which had been neglected for centuries. The earlier concept of LR, as we look back from today's planning perspective, was rather broad in that it not only sought to develop new urban land supplies but also emphasized the reshaping of the existing urban configurations.

Experiences in Kathmandu (Nepal)

LR, which is known as land pooling in Nepal, is still in the stage on infancy. Till 1999, eight land pooling projects have been completed and four are ongoing. All projects are in Kathmandu vally and the total area of the projects are 360 hectares in extent. These projects are backed by the Town Development Act and the act says that land pooling can be carried out in any part of the town planning area with the agreement of minimum 75% of the land owners (Nishani, 2003).

Experiences in Hawai'i

LR, when implemented with all its provisions, can be of assistance in the following situations, which may occur on O'ahu and the Neighborhood Islands: inner city district redevelopment, urban-rural fringe readjustment. LR as the potential to well address landowners, tenants and community's rights in urban redevelopment and rural LR. It also constitutes a way to finance infrastructure provision. The City and County of Honolulu General Plan gave priority directions to the redevelopment of Honolulu, before expansion at the urban fringe (Minerbi, 2002).

3. The redevelopment the MAS areas of Kampong Baru

Kampong Baru is located in the Central Planning Area of Kuala Lumpur within the proximity of commercial centres of the Golden Triangle to the east and the Central Business District of Kuala Lumpur to the west. There are many government and market development projects undertaken outside the MAS areas. Inside the MAS areas, however, no major development projects have been initiated except very few commercial development undertaken by private Malay developers. Figure 2 shows the physical characteristic of Kampong Baru area.

3.1. Land development initiatives in the MAS areas of Kampong Baru

Since being alienated in 1897, the MAS areas have been considered as an example to represent the Malay urban areas in Kuala Lumpur. The fact that the MAS areas have potential for development and, at the same time, are under-used reflects the condition of Malay landownership in Kuala Lumpur. There have been, development initiatives undertaken by various public and private agents to redevelop the MAS areas. However, the restriction in interests of the indigenous MAS land rights, together with other development problems, has led to the failure of these land development initiatives. As a result, the present state of the MAS areas is still under-utilised (Omar, 1999). In 1991, the government realized the importance of direct intervention to redevelop the MAS areas. The way proposed was to acquire land for redevelopment purposes and to plan a comprehensive and integrated land development scheme. In so doing, land development difficulties such as landownership and problems with landowners would be avoided.

3.2. The land development difficulties in the MAS areas of Kampong Baru

Although the strategies were designed for implementation to redevelop the MAS areas from 1991 to 1992, there were problems which led to the failure of the development initiatives. The main problem was high land prices as demanded by the affected landowners of the MAS land (Omar, 1999). High land prices asked by the affected landowners meant high land acquisition costs which affected the government decisions to undertake the redevelopment scheme. As a result of high compensation demanded by the landowners, in 1992, a statement leading to the cancellation of the Kampong Baru redevelopment projects was issued by the Prime Minister of Malaysia. As an alternative, another initiative was taken in 1995. A group of Malay corporate was formed to re-initiate a similar redevelopment programme. The group was established under the name of the Pembangunan Mas Melayu Sdn Bhd (hereafter known as PMMB). The redevelopment would have been undertaken in accordance with the Drafted Local Plan for Kampong Baru, and Kuala

Lumpur City Hall was once again appointed as the development coordinator (Kamaruzaman, 2006). The PMMB envisaged that the potential land redevelopment constraints would be in terms of physical and infrastructure problems, or of a legal and planning, economic and financial, socio-psychological and political nature (Omar, 1999). These problems were anticipated to include landownership, price and landowners' attitudes as forms of development constraints. Therefore, it was anticipated that the preferred way to develop the MAS areas would be as a package rather than a piecemeal approach. However, several attempts to acquire collectively contiguous sites for redevelopment were aborted due to the high prices demanded and a lack of participation from the affected multiple landowners. In this

failure point of view, another strategy that acquire individual sites through personal negotiation with the landowners with a view to redevelop the individual sites through the amalgamation of several contiguous sites. One of the strategies for redevelopment, therefore, was further used to negotiate with the landowners to buy individual sites or contiguous sites, wherever it was possible for amalgamation purposes. Although this strategy was a slow land transfer process through negotiation and agency social interactions, these strategy of the land development initiative may avoid landowners' tendency to expect higher prices for their lands. As shown in Table 1, there are several land development difficulties faced by various agents in their initiatives to redevelop the MAS areas.



Figure 2 The physical characteristic of Kampong Baru area

Table 1. The difficulties of land development in Kampong Baru

Period	Types of key agents involved	Development difficulties
During 1991-1992	Kampong Baru Redevelopment Committee (various property professional and consultants for development) MAS Board, Landowners	Limited market, lower values, valuation for compensation High expectation of prices Landowners' attitudes
During 1995-1999	Consortium Pembangunan MasMelayu Berhad (PMMB) & Malay corporates MAS Board and KLCH Landowners	Limited market, lower values, valuation for compensation High expectation of prices Landowners' attitudes
Field survey 2006	Consortium Pembangunan MasMelayu Berhad (PMMB) & Malay corporates KLCH Landowners	Multiple ownerships, High expectation of prices, Limited market, Landowners' attitudes

Source: Ismail, 1999 and field survey 2006

3.3. The LR more appropriate redevelopment for the MAS areas of Kampong Baru

Based on the experiences of other counties, constraints and the potential of the MAS areas, it shows that LR technique provides more appropriate approach for redevelopment this area. LR technique gives private landowner to actively participate in the development processes. Whilst, the government acts as land owners' partner that facilitates and controls the projects. In addition, the degree of successfulness of the project relies heavily on the land owners participation including the ability of the project to accommodate the community interest. The following discussion elaborates some problems which may be reduced by implementing LR method.

Multiple ownership:

Almost 99% lands in the MAS areas of Kampong Baru are multiple ownership. This is the main constraint to redevelop Kampong Baru area since disagreement from anyone of the land owners will influence the government programs. LR technique will handle this constraint by encouraging the land owners to actively participate in the project so that their interest can be accommodated precisely. As a result, the degree of disagreement and feeling lost can be reduced significantly. To achieve this objective, the land owners' knowledge about the LR method should be given prior to the projects.

Land price:

Another development constraint found in the MAS areas of Kampong Baru is the high land prices asked by the land owners. High land prices of course will hinder the second parties such as government or developers to undertake redevelopment particularly from economical consideration. It is due to every party try to get highest benefit from the projects. LR method threatens land owners as an agent of development. An other words, land owners don't have to sell their lands and developers just translate the landowner interest under government policy.

The land owners' interest:

Because land owners actively participate in the LR project, any interest such as whether the land owners want to occupy or to lease the new building or to maintain the old building can be easily accommodated. Including to the considerations is the willingness of the community to maintain the traditional building through replacement plan. There fore comprehensive redevelopment planning must be established and all of land owners must actively participate in the LR program.

Gradual and partial:

Implementation LR method for the MAS areas of Kampong Baru is different with other areas, because Kampong Baru is settlement area. Demolishing and rebuilding the area is absolutely expensive, so that gradual and partial of LR program is needed. Land owners can make a group based on location, interests or objectives to implement this program.

Preserving private land owner by Malay:

Since land owners don't have to transfer all their right by selling their lands, the Malay ownership over their lands still can be maintained. However as the consequences of the LR project, part of their land should be sold to support the redevelopment cost. To do that, restriction of land ownership for the MAS lands have to be considered.

Development partnership:

Redevelopment needs a huge amount of cost particularly to buy the lands and to redevelop the lands. By using LR method, the only cost to be paid is the redevelopment process. To do that, land owners can sell part of the lands to the Malay or to use part of the lands as a compensation to the developers. It means that in LR projects, the key process is development partnership which can be arranged by government.

4. Conclusion

The main aim of urban LR is to increase the supply of urban lands for development. To do that, fragmented and unstructured land parcels are amalgamated so that an integrated and structured planning can be designed. At the complex urban areas where the LR is hard to be implemented simultaneously for the whole area, the similar steps can be undertaken on a case-by-case basis either through voluntary conveyances or through enforcement (expropriation). In so doing, problem on property rights and ownership of the lands have to be handled before any buildings are constructed. Since MAS areas of Kampong Baru in Kuala Lumpur is a unique Malay residential in term of history governing it, the existence of this Kampong should be preserved. Thus, in order to keep balance to the development of surrounding areas, Kampong Baru need a reasonable project such as redevelopment or rearrangement approach on it. Based on aforementioned constraints, it seems that LR method is one alternative should be considered to redevelop Kampong Baru. To do that, socialization of LR method to all land owners in the MAS areas of Kampong Baru and other related agencies is the key step to gain the successful of the projects.

References

- [1] Adams, David, Disberry, Alan and Hutchison, Norman (1997) *Land ownership constraints within the development process*, Aberdeen Papers in Land Economy
- [2] Cunningham, Christopher Robert (2005) *Uncertainty, Zoning and Land Development*, A dissertation submitted for the Doctoral Degree in Economics in the Graduate School of Syracuse University
- [3] Doebele, William (2002) *Introductory Remarks*

Lincoln Institute of Land Policy, Conference Paper

- [4] Hamlin, Roger E. (2003) *Public-Private Partnerships for Inner-City Redevelopment*, Institut for Public Policy and Social Research and Department of Geography, College of Social Science, Michigan State University
- [5] Hayashi, Kiyotaka (2002) *LR as a Crucial Tool for Urban Development*, Lincoln Institute of Land Policy, Conference Paper
- [6] Minerbi, Luciano (2002) *Efforts Toward LR Legislation in Hawai'I*, Lincoln Institute of Land Policy, Conference Paper
- [7] Nishani, Wickramaarachchi (2003) *LR as a Land Management Techniques for Urban Areas*, 9th International conference on Sri Lanka Studies, November 28th – 30th, Matara, Sri Lanka
- [8] Omar, Ismail (1999) *The Supply Constraints on Indigenous Lands in Kuala Lumpur*, A thesis submitted for the Doctoral Degree in Land Economy at the University of Aberdeen.
- [9] Payne, Geoffrey (2003) *Building Bridges: The Experience of Public-Private Partnerships in Urban Land Development*, Lincoln Institute of Land Policy Conference Paper
- [10] Sagalyn, Lynne B. (2002) *Land Assembly, LR and Public/Private Redevelopment*, Lincoln Institute of Land Policy, Conference Paper
- [11] Viitanen, Kauko (2002) *The Finnish Urban LR Procedure in an International Context: What Can Be Learned Commonly*, Lincoln Institute of Land Policy, Conference Paper

Transformation: Impact of the Federal Capital Development on Gbagyi (Gwari) Indigenous Housing pattern, Abuja, Nigeria, 1976- 2005

Mai, M.M.¹, Shamsuddin, S.²

Dept. of Architecture,

¹Faculty of Built Environment Universiti Teknologi Malaysia
Tel. +60-146262418, e-mail: mouktarmai@yahoo.com

²Dept. of Civil Engineering, College of Science and Technology
UTM City Campus, KL, tel. 03-26154872, e mail: shuhana@citycampus.utm.my

Abstract

Housing pattern and form are partly the result of human adaptation to ecology and man's attempt to make and express meaning out of the environment nonverbally. In the course of realizing this, however, human conceptions of the self and the resultant behavior, as well as factors of social production and reproduction come to play. It is common knowledge that conceptions of the self and or, factors of social production and reproduction vary from one environment to another. This implies that, meanings people give to an environmental setting is relative to the space, time and context. Although societies are by nature dynamic, these changes take place gradually over long periods of time and thus they are not normally noticeable except in case of great social and political interruptions, such as the decision to make Abuja Nigeria's capital. Though the pronouncement to relocate the seat of Government from Lagos to Abuja was made almost three decades ago, it only assumed the capital status fully for less than a decade and a half by 2005. Abuja's new political position culminated in significant changes in its built environment in general and its indigenous traditional housing pattern in particular. Such housing adjustments invariably affected the settlement, with its attendant socio- economic implications. The cumulative sum of all these changes was the lost of spatial identity of the host, Gbagyi community especially in the municipality. This study evaluates the effects of the FCT development in transforming the indigenous housing pattern of the host community. Housing transformation is today a key source of alternative housing supply in most urban centers of developing countries. As such, the significance of this study is to identify the factors leading to the transformation of indigenous to vernacular housing pattern incorporating multi- cultural dimensions.

Key words: transformation, housing pattern, human behavior, spatial identity, culture

1. Introduction

Housing environment is influenced by changes in economic forces of demand and supply, socio-cultural factors like family's basic needs and home ownership, regional demographic imbalance such as population shifts and affordability. The manifestation of these housing transformations could be in the form of spatial pattern, building appearance (physiognomy), and housing occupation. The sum total of the various housing transformation is an adjustment in the settlement pattern. The implication of these includes: a distortion of the planning scheme; pressure on existing infrastructure and amenities; and culture change of the immediate community, (Nguluma, 2003: 84-100).

The aim of this paper is to highlight the effects of the federal capital development on the host Gbagyi community's indigenous housing pattern, especially the acceleration of housing transformation occasioned by the relocation of the presidency from Lagos to Abuja. The discussion will focus on three significant stages of the development of the new federal capital territory, FTC-Abuja (Junaidu, 2004:134-161). These are: the incipient (1967-1987); the intermediate (1988-1992); and the consolidated (1993-2005). These will form the basis of further discussions in six sections. First is an over view of the

birth of FCT-Abuja. Second is the theoretical framework of the study. Third is the research approach and methodology, which involves a mix of quantitative and qualitative paradigm. Fourth is the presentation of results relevant to the study. Fifth is the discussion of findings and finally comes the conclusion.

1.1 The Birth of FCT- Abuja

Abuja was declared Nigeria's new federal capital in February 1976. Its development was implemented in phases by successive regimes over a period of three decades so far. A systematic and dramatic change in the social and political fabric of the Gbagyi land now referred to as the Federal Capital Territory, FCT-Abuja occurred as soon as the presidency moved to Abuja 1991. Population shifts from other parts of the country posed a threat to the survival of the indigenous Gbagyi housing style.

Socio-economic changes spontaneously set in as a response to the economic forces of demand and supply; thereby changing the psyche of the general

populace at the personal and individual level; as well as the communal and societal levels; cumulatively. This results in modified and in many cases, new conceptions of the self, identity, cosmology and eschatology. Therefore, there were behavioral and attitudinal changes that became manifested in factors of social production and reproduction (Oumar, 2004:3). The changes brought about by the relocation of the federal capital to Abuja led to substantial and important changes in the built environment, in particular the house and the settlement not only in Gbagyi land but also in adjoining settlements of the neighboring states. Some of these changes had social and economic implications. Castells (1997:6) in Dogara (2005:14) maintains that "identity is a people's source of meaning and experience."

- Incipient Stage (1976-1987): this is the early preparation period during which professional inputs from architects, engineers, and planners were made; public policies enacted; and Gbagyi settlements played hosts to other Nigerian settlers, most of whom were construction workers, traders and other business concerns. Demographic changes were slow and steady; Gbagyi houses accommodated most of the settlers first as tenants.
- Intermediate stage (1988-1992): the Federal Government raised the tempo of the construction of infrastructure in FCT Abuja, as such Nigerian job seekers and businessmen responded accordingly via a high population shift from other parts of the country to this new found land. More importantly, the presidency relocated from Lagos, thereby sending a clear signal to all and sundry that Abuja is a reality. Public and private concerns now trooped to the FCT in search of office and residential accommodations. This scenario impacted on the Gbagyi settlements. Housing extensions, adjustments, and new construction became the order of the day. Farmlands were subdivided into residential plots; uncompleted buildings were occupied by unknown fellows; and public services like water and electricity were overstretched.
- Consolidated stage (1993-2005): the year 1994 witnessed the final leg of the movement of Federal Government agencies from Lagos to Abuja. This also marked the elastic limit of Gbagyi settlements to accommodate in the trooping Nigerians. At the same it coincides with the return to democratic dispensation in the country in February 1999. Residential and commercial developments now took a "*fill in the gap strategy*," (Tippie, 1996). These entail extending rooms to cover setbacks; building on courtyards; and converting residential rooms adjacent to roads into shops.

2.0 Theoretical Framework

Vernacular housing in most cities in the developing countries is currently in transition, (Oliver, 1997: ix-xi) due to the non-equilibrium between demand and supply. Housing deficit is the order of the day. Tripple et al (2000:3) declared that recent expansion of cities in sub-Saharan Africa has tended to be in the form of squatter invasions and informal subdivisions or of sprawling suburbs of medium- to high-cost housing. In most

cases, they spread outward and fill in spaces between villages subsumed within cities' built-up areas. Such is the state of affairs in Abuja, Nigeria, the subject of this discourse.

Traditional Gbagyi settlements in FCT-Abuja are characterized by three housing patterns. These include: homestead, farmstead, nuclear and extended family compounds in nucleated and ribbon settlements, (Balogun, 2001:99). These spatial housing patterns are manifested in two typologies: Sudanese round huts and/or west coast rectangular type, (Ibid: 100). These houses are normally produced informally through self help and communal effort. A father usually erects a core housing unit (bedroom and Kitchen) for his son in the family compound or elsewhere, to start his own family. Formal housing delivery is hardly practiced in Gbagyi culture.

Informal housing in Nigeria refers to agency or government provision of completed housing units which may be allocated or acquired by the citizenry. It could also be in the form of official land acquisition followed by approved building construction by any means. Informal housing is also referred to as popular housing in developing countries, because it is a common access to shelter in such economies. Since the informal or self help housing is readily practiced in the Third World, the application of simple technology, and the use of local materials are inevitable. No wonder therefore, mud round huts roofed with thatch are typical rural housing typology. Such rural norms are combined with rectilinear spatial pattern at the semi-urban level. However, the urban housing, especially in peri-urban Abuja is characterized by urban norms, whereby zinc or aluminum roofs on mud or concrete structures are being used. Nevertheless such durable houses are produced through informal mode, incrementally, at the economic convenience of the self help developer. Factors influencing house form adjustments could be: socio-cultural needs like the seclusion of women from male visitors (privacy); income generation from renting out some spaces; creation of space for home businesses; livestock sheds; etc. Each of these space requirements will affect the existing housing pattern in terms of nature of occupancy, spatial configuration, and environmental quality (Legorreta, 1984:9) in (Walker, 2001:6).

House form may draw from an existing cultural setting and social values of its people but economic factors exert considerable influence over housing development and transformation. Analyses of the changing mode of socio-economic set up within a period and the spatial expression of division of labor are, thus vital in the evolution and changes in housing forms, (King in Chokor, 1986). An equally significant factor in housing design is the fact that housing as organized physical forms, are settings where people have deep personal experiences with

profound social meanings, and symbolizes individual and group values, (Chokor, 2005:70).

3.0 Research Approach and Methodology

The objectives of the research are: to identify Gbagyi indigenous housing pattern; transformation of the pattern over time; and motivations for the transformation. The data required for these are quantitative data that will indicate trends and patterns of housing transformation socially and physically. Then qualitative interviews and literature review will aid in triangulation and validation. Mix quantitative and qualitative research paradigm is adopted for an enhanced understanding of the process of housing pattern transformation, which demands an exploratory descriptive, and explanation of the phenomenon. Research instruments used are: questionnaire (210 random samples); key informant interview, (14 respondents); in-depth interview, (21 respondents); and scaled drawings, (21 sets), to act as a means of comparison with the social survey data. Analysis is done in three levels. First, the quantitative or statistical information is appraised to establish a trend or pattern. Then the social meaning of the theme is searched within the key informant views. These will now be compared with the key informants' responses.

4.0 Results

The physical and social structures of Gbagyi settlements could be analyzed in time series with focus on the three stages of development. Incipient (1976-1987): the social composition was 70% Gbagyi and 30% other Nigerian settlers. As for house form, 95% mud round huts called Sudanese housing type; and 5% rectangular mud West coast housing type. As at that time Gbagyi preferred developing houses on family land. This is due to the value attached to the power of the ancestral blessings symbolized by their paved graves within the family compounds. This period coincided with the resettlement of the Gbagyi community within and outside FCT-Abuja to pave way for smooth development of public infrastructure.

Intermediate stage (1988-1992): Gbagyi population decreased considerably due to formal and informal resettlement of the indigenous. A multi-ethnic population increased considerably in the study area, due to employment opportunities offered by the massive construction work. As such a homogenous society was gradually developing in peri-urban Abuja. The rural round huts were being systematically replaced by rectangular housing pattern. Balogun (2001:100) opined that the Sudanese round huts reduced to only 20%. The Nigerian settlers were busy acquiring plots of land from their Gbagyi hosts. Gradually farmlands were subdivided into residential plots and disposed off settlers and absentee land speculators. Gbagyi people lost their emotional attachment to ancestral graves. Instead, value for monetary gains became more pronounced. The FCT administration banned home graves with public cemeteries being developed at various settlements.

Consolidated stage (1993-2005): A socially heterogeneous urban setting with high population density was developed in Karu and Nyannya villages. Jikwoyi maintains a medium population

density. Urban norms replaced rural ones especially Karu and Nyannya villages fully. New housing typologies of face to face rental housing, and owner/tenants occupancy evolved. Most of the rental housing units were constructed on former Gbagyi farmlands. Public cemeteries were consolidated and community leaders are entrusted to assist in the patronage of public cemeteries.

The study shows that motivations for transformation vary from socio-cultural requirements to economics of income generation. 59% of the 210 samples extended and/or adjusted their houses to earn income from rent; 10% were driven by the need for home income generation; and 31% by increase in family size. These transformations coincide with Kelett's (2002) view of home construction. Kellett (2002:1) asserted that an informal settlement is an expression of the complex interrelationships between dwellers and their dwelling places. This is based on culturally constructed images of what dwellings and settlements should be like. Indeed the process of incremental growth 'imaged futures' (Holston 1991) in (Kellett, 2002:1). However, Walker (2002:3) argues that informally produced housing units are improved relative to socio-cultural values which are constantly renegotiated. Such an approach demands a clear comprehension of the relationship between house form and the underlying social and cultural processes.

5.0 Discussion

The fieldwork survey of this study indicates a metamorphosis of the Gbagyi social and physical structures. The high homogeneity of Gbagyi settlements and residential occupations in peri-urban Abuja reduced with the advent of the relocation of the presidency from Lagos at the end of 1991, as people of different ethnic groups inhabit the same environment. It could be observed that most of these inhabitants belong to the low and middle income groups. As at the first quarter of 2005, 46% of the inhabitants of Nyannya, Karu and Jikwoyi informal settlements are from Northern Nigeria (Hausa-Fulani, Gbagyi, etc); 26% from South East (Ibo and similar ethnic groups) and 28% from the South West (Yoruba and its derivatives). This implies that the Gbagyi indigenous inhabitants have been reduced from 80% as late 1970s to about 25% of the population in 2005. Some of the physical features of Gbagyi indigenous housing pattern are enumerated below in three periods:-

Incipient Stage (1976-1986) is physically characterized by the family head's granary located at the center of the house, while others are sited outside; Tombs or paved graves are commonly found within the courtyard of rural settlements; Parking of logs of wood is a common sight in front or at the back the compound, which suggests the liquidity or economic well being of the women in

the house. A traditional grinding mill called the grinding stone, functions as the corn mill for a complete neighborhood, where the processing of corn to flour is manually carried out by women communally. This could not be sighted anywhere in the study area, though enumerated by a number of key informants. The presence of sacred rooms for idol worship and associated rituals is typical of pagan Gbagyi housing typology, (Garba, 1990: 87). Multiple exits in the house is another feature of rural Gbagyi housing which extends to semi urban and urban settings, similar to Yoruba urban housing style observed by the authors in Lagos, Nigeria. Spaces within dwelling units flow into one another such that the physical the boundaries between neighboring houses is better imagined than real, as remarked by Denyer (1978:19). This exhibits loose hierarchy of spaces of different ownerships especially in rural settlements. The absence of toilets in the Gbagyi dwelling is customary in rural areas.

Intermediate Stage (1987-1992): As the settler population increased, demand for tenement housing exceeded supply. Consequently, the Gbagyi indigenous population was persuaded to admit tenants into their dwellings. This gave rise to owner/tenant occupancy and its attendant housing transformations. Mix curvilinear and rectilinear compounds in which most of the Sudanese round huts were replaced with West coast type due largely to the influence of settlers; and improved income of Gbagyi peasants. Such incomes come from land deals, blue collar jobs and house rent. As some Gbagyi youths were engaged in the construction industry and associated activities, the traditional extended family system got weakened.

Consolidated (1993-2005): Highly heterogeneous community now inhabits peri-urban Abuja settlements. Gbagyi population has considerably reduced due to voluntary and institutional resettlement of the 1980s. High population density due largely to housing shortages, especially for the urban poor, becomes synonymous with Abuja. Thus, high housing occupancy rate and shared accommodation was considered a feasible option (Ikejiofor, 1998). This housing confusion resulted in the development of a variety of housing types such as face to face rooming accommodation; compound housing; and flats. All these adopted a variety of spatial identities in rectilinear pattern.

5.0 Conclusion

Undoubtedly, the nature and form of a housing pattern is a function of human efforts to adapt to the ecological setting and his attempt to make meaning out of the environment. Consequently, shifts in human perception and or condition invariably lead to transformation of housing units. The relocation of the presidency to Abuja in December 1991 was primarily a social impetus which led to substantial and important changes in the Gbagyi indigenous housing pattern in FCT- Abuja in general. The paper looked at the changes in the physical aspects of the settlement and the socio-cultural changes in the domestic house. The authors thus try to abstract the impact of the changes brought about by the population

shifts occasioned by the new status of Abuja in the economic morphology of Nigeria.

The housing pattern at the incipient stage of Abuja had three main characteristics, which were greatly modified as a result of the population explosion. Firstly, the normal organic nature of growth and development was reinforced with a state policy of establishing new resettlements sites in and around the FCT, and upgrading existing spontaneous settlements. Secondly, the spatial morphology of the settlement, both in terms of form and function was modified to fit in with the revived economics of demand and supply. Finally, new impetus in the sense of certain social incentives was given to the hitherto slow-moving demographic dynamics.

The Gbagyi house has had a long history of evolution, starting from a simple collection of basic habitations encompassed by a wall or mat reed fence joining rooms, called house, into the complex organic whole, called *compound*. This gradation is closely related to the socio-political evolution of the Gbagyi society. The rapid urbanization of Abuja facilitated the adaptation and spread of the *functional rental houses* (face to face type) over and above the *multi-generation compound* particularly in the urban areas. In addition, it modified the urban housing norms such as the avoidance of organic building materials; more internal subdivision of space; and access to formal services like water and electricity.

The motivations for these housing transformations includes: home construction through income generation at home, (Kellet, 2002); socio-cultural demands of privacy- seclusion of family, (Rapoport, 1969); and the house as communication of social identity of its inhabitants as typified by Karu Hausa settlement.

References

- [1] Balogun, O. (2001): **The Federal Capital Territory of Nigeria: A Geography of its Development**; Ibadan University Press, Publishing House, University of Ibadan, Nigeria.
- [2] Chokor, B.A (2005): **Changing Urban Housing Form and Organization in Nigeria: Lessons for Community Planning**, Journal of Planning perspectives, 20, January 2005, pp70
- [3] Denyer, S. (1978): **African Traditional Architecture**: Africana Publishing Company, New York
- [4] Garba, B. T: **Garki Village Resettlement Scheme Abuja: An Exposition of Gbagyi (Gwari) Traditional Architecture**, Unpublished Master of Science in Architecture Thesis

- (1990), Ahmadu Bello University, Zaria, Nigeria.
- [5] Gwamna, D. J. (Ed): **Gbagyi Journal**, Vol. 2. 2005, Gbagyi Vision Publications (GVP), pp14; 2002: 43.
- [6] Ikejiofor, U. (1998): **Tyranny of inappropriate policies: Sharing as housing strategy among middle/low-income households in Abuja, Nigeria**, *Cities*, Vol. 15, No. 6, pp. 429–436, 1998
- [7] Junaidu, A. M (2004): **Analysis of Urban Influence on Rural Settlement Growth and Land Use Cover in The federal Capital Territory of Nigeria Using GIS Data**. An unpublished PhD thesis, FUT Minna, Nigeria
- [8] Nguluma, H. M (2003): **Housing Themselves: Transformations, Modernization and Spatial Qualities in Informal Settlements in Dar es Salaam, Tanzania**, Department of Infrastructure, Royal Institute of Technology, Stockholm, unpublished PhD thesis.
- [9] Oumar, A.A.M (2004): **JUYIN ZAMANI: Impact of Sakkwato Jihad on the Kano Built Environment**. *Proceedings of the Conference marking the 200 Years of the Usman D'an Fodio Jihad in Kano*, p p3.
- [10] Oliver, P. (ed.) (1997): **Vernacular Architecture**. Published by Barrie & Jenkins Ltd. London
- [11] Rapoport, A. (1969): **House form and Culture**, Prentice-Hall, Inc. Eaglewood Cliff, N.J.
- [12] **The Gbagyi and the National Question: A Profile of Marginalization and Demand for Justice**. *A Submission by The Gbagyi of Nigeria, to National Political Reform Conference*, March 2005:3
- [13] Tipple, G (1996): **Housing Extensions as Suitable Development**, HABITAT INTL. Vol. 20 No.3, 1996 pp376-376
- [14] *Tripple et al (2000): User-Initiated Extensions in Government-Built Estates in Ghana and Zimbabwe: Unconventional but Effective Housing Supply*. A. Graham Tipple, with Stephen E.Owusu and Columbus Pritchard Africa Today, pp 80
- [15] Walker, A. (2001): **The Social Context of Built Form: The Case of Informal Housing Production in Mexico City**. Working paper No. 114, Development Planning Unit, University College London, pp 3, 6

Development of Vegetation and Land Use By Remote Sensing Method in Wadi Tuban Abyan Area - Yemen

A. A. Almhah

Department of remote sensing
Faculty of Geoinformation Science & Eng
University Teknologi Malaysia
ayoub.almhah@fkg.utm.my

I. Busu

Department of remote sensing
Faculty of Geoinformation Science & Eng
University Teknologi Malaysia
ibusu@fkg.utm.my

N. Ibrahim

Head Department of Geoinformation
Faculty of Geoinformation Science & Eng
University Teknologi Malaysia
nourkhair@fkg.utm.my

Abstract

As part of the Water Resources Management Study of the Wadi Tuban and Wadi Abyan areas in the Republic of Yemen, reviewe on existing geomorphological, crop patterns, geology and land-use data, and constructed set of summary maps has been carried out. The existing geological and land-use mapping is improved using modern satellite image interpretation. This paper describes the surveys, results and also provides a comprehensive interpretation of all of the reliable data. The results of the geological mapping can be found in the paper. The study area cover two regions of approximately 1500 km² along the southwestern coastal area of Yemen in proximity to the city of Aden. Remote sensing and geographical information system have been creatively utilized to assist in the study and better understanding of the spatial variation and distribution of certain processes in vegetation, land use and water resources . Hydroinformatics is the contemporary term that best describes the interactive knowledge-based non-expert decision support systems (DSSs) which comprises of information system and a number of databases.

1. Introduction

The study of Water Resource Management in the Tuban and Abyan Deltas in the Republic of Yemen the geomatics was used to conduct a remote sensing and mapping study with the following activities:

Acquire and interpret current satellite imagery for land-use and geological information;

- Gather existing land-use and cultural data for the Wadi Tuban and Abyan regions;
- Gather existing surficial geological and soils data for the Wadi Tuban and Abyan regions;
- Describe and map dynamic land-use elements relevant to the water resources study; and
- Construct a digital elevation model from available topographic information for the study area.

The planned 1:100,000 scale land-use and geoscience maps were split into two sheets covering the Tuban and Abyan Deltas respectively. The types of maps involved are:

- The basemap based on the Landsat TM scene;
- The topographic map;
- The soils and geomorphology map;
- The dynamic land-use and vegetation maps.

In addition to these maps, a 1:250,000 scale basemap was prepared to provide assistance .



Figure 1: The location of area of study

2. SATELLITE IMAGES

The satellite images used in this investigation comprised:

- Landsat TM scenes acquired on August 21, 2000, Processing of the satellite imagery included:
- Image registration and georeferencing;
- Re-projection into UTM, Zone 38N, WGS84 coordinate system;
- Filtering and enhancing image bands for geological, surficial and vegetation interpretation;
- Combining Landsat TM bands 7,4,1 as Red-Green-Blue (RGB) for use as a mapping base; and,
- Merging with digital interpreted basemap features.

All of the LandSat channels from both scenes will be provided to National Water Resources Authority (NWRA) in the format received from RadarSat, Inc., including appropriate geo-referencing information.

The LandSat TM imagery includes 8 bands of spectral data, each band with a distinct range of wavelengths. The resolution for most of the bands (1-5, 7) is 28.5 m x 28.5 m. Band 6 has a 60 m resolution and band 8 has a 15 m resolution.

3. Groundtruthing

field checking in both the Tuban and Abyan Deltas has been conducted. Because it is difficult to know what access and actual ground conditions are, only the types of locations required were identified.

For each location, the following information was collected:

- Location Identifier
- UTM coordinates (collected with a GPS unit)
- Description of Picture Location
- Date
- Description of the Vegetation Type and Cover
- Description of the Irrigation Type
- Description of the Terrain Type
- Description of the Surface structure
- Description of the Population
- Notes

The field checking was used to confirm vegetation types and populated area interpretation. The results of the groundtruthing can be found in Figure 1.

4. Remote Sensing Interpretation And Data Collection

A number of data layers were required as part of the terms of reference to be interpreted from the satellite imagery and other data sources. Our primary method for this data collection was to acquire and review all available and existing maps, evaluate these maps for their usefulness and then extract the required information. This information was then compared against the current satellite imagery and updated as required.



Figure 2: field tests points location map

The major data layers collected/created are as follows:

- Drainage Features
- Towns and Cities
- Historical and Archaeological Sites
- Topography
- Wadi Floor Material Type and Soil Quality
- Current and Historical Vegetation Patterns
- Current and Historical Population Patterns

4.1 Drainage Features

Drainage features were collected in a similar manner as the transportation features. The wadi channels were digitized originally from the 1977 topographic basemaps and then updated based on the 2000 LandSat TM imagery.

Braided wadi channels were shown where it was seen as appropriate.

4.2 Towns And Cities

The locations of towns and cities were collected from the 1977 topograph base maps. The names of these locations were translated by National Water Resources Authority (NWRA) staff and provided to Komex to be used as the standard English spelling for all maps and figures. Towns and cities were differentiated by the size of the population.

4.3 Topography

Topographic data for the detailed study area was required so as to produce 20 m contours and to provide a digital elevation model for future water resource studies.

The contours for the Tuban and Abyan Deltas were produced using the 1:100,000 scale Russian topographic base maps and digitizing the contour

lines. These digital contour lines were then used in conjunction with the spot elevations available on the topographic maps to generate a three dimensional model of the study area. Once the three dimensional model was complete and verified for accuracy, the 20 metre contours were generated using a modelling package called Vertical Mapper. These contours were then exported to ArcView for mapping.

Once in ArcView, the contour lines were plotted on the LandSat TM image and reviewed to ensure that they followed the topographic surface of the deltas and were cartographically correct.

The elevation data was collected based on mean sea level values referenced to the Hodeidah datum. The results of this work are shown on Map 3 (a) and (b).

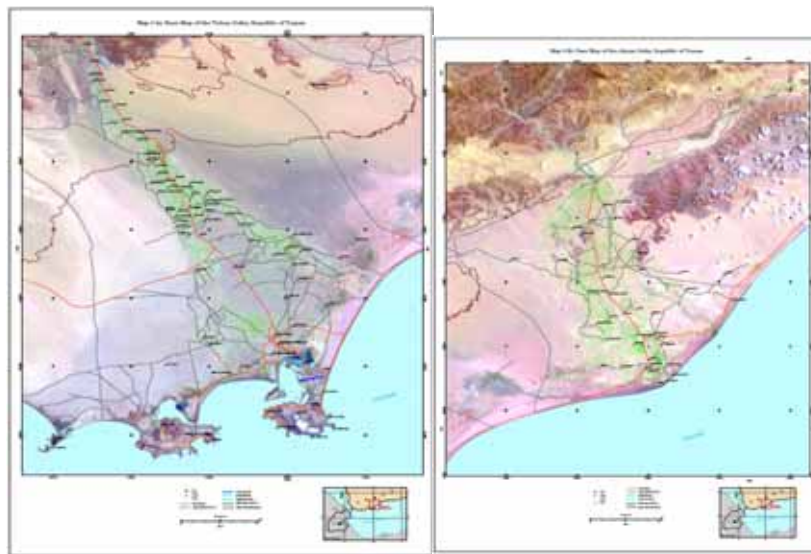


Figure 3(a) and (b): Topographic maps.

4.4 Vegetation And Cropping Patterns

Vegetation and cropping pattern data was collected over two different epochs. It is in the year of 1977 and 2000.

4.4.1 VEGETATION AND CROPPING PATTERNS IN 2000

Vegetation and cropping patterns in 2000 were based on the LandSat TM image. Two different methods were utilized to classify the vegetation in the Wadi's Tuban and Abyan.

Another common method of vegetation classification is called the Normalized Difference Vegetation Index (NDVI). This procedure highlights the differences between the infrared bands and the visible red bands and serves as a measure of vegetation health and water content.

Bands 3(Red light) and 4(near-infrared) were used for this calculation.

The formula $(B4-B3)/(B4+B3)$ was used. This

equation gives a range of values between -1 and 1. Upon inspection of the histogram and natural break analysis, the values can be grouped into categories that show ranges of vegetation health.

The output from this classification was used as input into the final classification scheme.

Supervised Classification

The results from the previous classification was used to verify the results from this classification. In addition, the groundtruthing survey was very important here to identify specific crop types that were present in any particular area.

In the supervised classification, training areas were selected based on visual inspection and reference to the ground truthing locations to represent categories of crop types. The spectral signatures of these crop types were calculated and applied to the rest of the image, resulting in the final classification as shown in figure 5 a and b.

For the Wadi Tuban, 4 categories were identified. These included:

Dense vegetation of more than 60 % ground cover including dura, seif, sesame, date, wheat, papaya, banana and dense natural trees & bushes

Light vegetation of 30-60% ground cover including natural grass and bushes, cotton, sesame, peanuts and various fruit crops.

Very light vegetation of less than 30% ground cover including natural grass and bushes and some light crops

Fallow - lands that are not currently vegetated.

For the Wadi Abyan, 3 categories were identified. These included

Dense vegetation of over 60 % ground cover including dura, seif, sesame, date, wheat, papaya, banana and dense natural trees & bushes

Light vegetation of 10-60% ground cover including natural grass and bushes, cotton, sesame, peanuts and various fruit crops.

Fallow - less than 10% ground cover - lands that are not currently vegetated.

4.4.2 Vegetation and Cropping patterns – 1977- 1996

As we did not have any satellite imagery for this epoch, existing maps were used to identify areas of vegetation and crops. The primary source of information was the 1977 topographic base maps.

These maps were incorporated as accurately as possible and presented in figure 4a and b

4.5 POPULATION PATTERNS

4.5.1 Population Patterns - 2000

Population patterns for 2000 were primarily based on visual interpretation of the LandSat TM image. In a number of areas, it was difficult to conduct a classification for the populated areas because the main feature that differentiates these areas from the surrounding wadi material is not color, but texture and patterns. Computer algorithms for pattern detection are not as reliable or well developed as algorithms for spectral signature or color detection.

Using the town and city locations from the 1977 topographic basemaps as a guide, the image was analyzed at a scale of 1:50,000, twice the scale of the required output, for populated areas. Regions that

demonstrated the typical texture and pattern of a built up area were traced on the image and associated with a town or city name. Because the pixel size of the LandSat TM image is 28.5 m x 28.5 m, some small regions were unidentifiable if the houses were small or there were very few houses. In these cases, band 8 of the LandSat TM image was used as it has a spatial resolution of 15 m x 15 m.

The results of this interpretation can be found in Figure 5 a and b.

4.5.2 POPULATION PATTERNS – 1977-1996

Populated areas for 1977/79 were primarily based on the 1977 topographic maps. These maps were geo-referenced and the areas representing built up and populated areas were digitized to provide the spatial distribution of populated areas..

The results of this interpretation can be found in Figure 5 a and b.

5- THE EXPECTED RESULTS

Understanding of the vegetation and land use growth process is highly crucial to vegetation and land use development planning specially irrigation areas expanded in the mean basin which have water shortage. Thus, modeling land use growth, production can be highlight useful for exploring the interaction between the Limited Nation Resources and built environment, to help national resources planning regarding decision-making complexity.

Under transition rule, multiple data source and factor are incorporated into determining the state changes. The result (after finish the theses) is a spatial representation of suitability for all cells to be converted to vegetation and land use for a particular time cycle. To create realistic and predictive vegetation growth and land use by time series, growth rate factor are calculated by the annual growth rate.

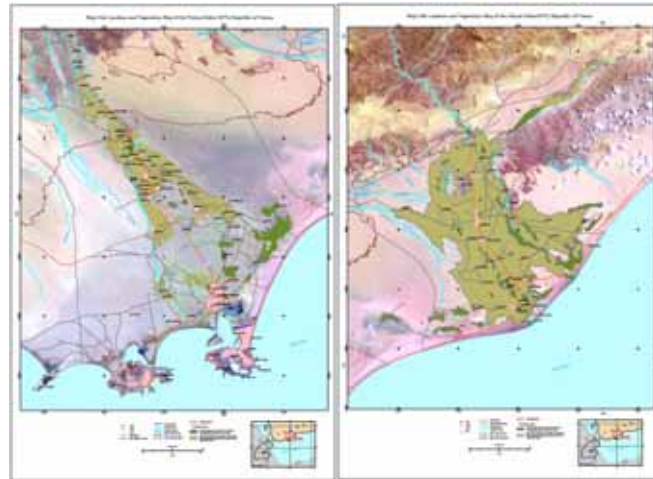


Figure 4 (a) and (b): Land use and vegetation map

6. Conclusions

A remote sensing and geographical information system have been creatively utilized to assist in the study and better understanding of the spatial variation and distribution of certain processes in vegetation a, land use and water resources . Hydroinformatics is the contemporary term that best describes the interactive knowledge-based non-expert decision support systems (DSSs) which comprises of information system and a number of databases.

5. References

- Beydoun, Z. R., 1997. Introduction to the revised Mesozoic stratigraphy and nomenclature for Yemen. *Petroleum Geology*, 14, 6, 617 - 629.
- Komex International Ltd., 1997. Groundwater Resources Assessment, Tuban- Abyan Region, Republic of Yemen. Volume I Text. Unpublished Report Prepared for Ministry of Oil and Mineral Resources, Republic of Yemen, November, 1997. Report No. KI94-3866 T18.
- Komex International Ltd., 2002 Draft Final Report – Tuban- Abyan water resources study , National Resources Authority, Yemen. March 2002.
- Topographic Base Maps, D-38, 39, 102, 103, 114, and 115, 1:100,000, 1977

Impacts of Electronic Data Interchange (EDI) Benefits on Firm Performance: Evidence from Malaysian Manufacturers

Gengeswari, K., ^{1*} and Abu Bakar, A.H. ²

Department of Management,
Faculty of Management and Human Resource Development
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
¹Tel: +6075531899, Fax: +6075531908, Email: k.gengeswari@gmail.com
²Tel: +6075531903, Fax: +6075531908, Email: m-abakar@utm.my

Abstract

Implementation of electronic data interchange (EDI) has been a long-standing issue among IT/IS scholars. Scholars have reported a constructive result towards success EDI implementation where EDI has been viewed as vital for transmission of business data. Thus, this paper is intended to further examine the implementation of EDI among Malaysian manufacturers. This paper examines the impact of EDI benefits on overall firm's performance as well as the extent of EDI integration and factors that influence EDI implementation. Mail surveys from 71 manufacturers were the main data collection method for this study. Findings indicate that the extent of EDI integration among surveyed respondents is just average. Top management support, sufficient financial resources and imposition by government were identified as significant factors towards EDI implementations. It also demonstrates that EDI benefits have significant impacts on overall firm performance.

Key Words: Information Technology and Electronic Data Interchange

1. Introduction

Electronic data interchange (EDI) is often referred as a type of interorganizational IT that enables exchange of business data with little or without human intervention between information systems of trading partners based on an agreed standard format. EDI was initiated in United States by transportation companies when they began to standardize documents and replace paper-based communication method with an electronic system [1]. Yet, the use of EDI became popular in late 1980s and early 1990s. To date there are more than 300,000 EDI users worldwide where its market was forecasted to raise considerably at the growth rate of 200% per annum [2]. Further, value of EDI software, products and consulting services was also predicted to increase from \$800 million in 1997 to \$2 billion annually in 2001 [3]

Applications of EDI range from as simple as transmission of purchase order and also as sophisticated as electronic fund transfer. EDI could yield enormous benefits provided it is integrated with other business applications. Otherwise, EDI could yield very minimal benefits where at this extent EDI would act a little better than electronic mail [4]. Several studies have demonstrated attainment of EDI benefits subject to the extent of EDI integration with internal business applications as well as with trading partners' applications [5] [6] [7]. The cooperative nature of EDI signals that EDI is a supplementary tool instead of substitute tool for the existing business applications. Benefits from implementation of EDI shall be used to improve overall firm performance particularly in the context of several functional areas including shipping, accounting and purchasing. Enhanced EDI benefits resulting from extensive integration with other applications could change the way of doing entire business. Many studies have confirmed the positive impacts of EDI benefits on overall firm performance [5] [8] [9].

* Corresponding Author. Email:
k.gengeswari@gmail.com, Tel: +6075531899, Fax:
+6075531908

Despite the substantial benefits and positive impacts on firm performance, many companies are reluctant to further

implement EDI or often stall with a single set of transactions [2] [10]. In Malaysia, for instant, the major motivation behind EDI adoption among companies is due to enforcement by government instead of self-initiatives [11] [12]. Companies that engage in import/ export activities are mandated to implement EDI for the sake of customs declarations. This scenario is applicable for many countries including Hong Kong, Korea, Singapore, Brunei and Taiwan. Resistance towards extensive EDI implementation at Malaysia is due to Malaysians' unwillingness to change from their old values or from using conventional trade methods [13]. Besides, lack of internal technical persons, insufficient financial resources and unsatisfactory top management support have been identified as the major inhibitors for EDI implementation.

Yet this justification can be too hasty as there is only little past EDI literatures within Malaysia context [11] [12] [14] [15]. Thus, more empirical researches are required to further examine the EDI implementation at Malaysia. Compared to Malaysia and even South East Asia, majority of past EDI studies were confined to the USA, Canada, Europe and Australia. By taking all these facts into account, this paper is intended to examine EDI implementation in Malaysia by focusing on manufacturers. The remainder of this paper as follows: next section reviews briefly the related literatures; third section discusses how this study was conducted; fourth section, collected data was analyzed and discussed; and this paper ends with brief yet concrete conclusions.

2. Literature Review

Consistent with the growing importance, subject of EDI has received special attention by scholars and practitioners. EDI researches cover several similar topics such as EDI benefits [2], drivers [7], inhibitors [16] and extent of integration [8]. Very less studies have researched impacts of EDI on firm performance yet exception can be given to the [8] [17] [18]. Thus, as stated before, this study examines the extent of EDI integration, key factors that influence EDI implementation as well as the impacts of EDI benefits on overall firm performance.

2.1 Determinants

Iacovou et al. [7] found three EDI adoption factors namely organizational readiness, external influences and perceived benefits. Chwelos et al. [3] modified the findings of Iacovou et al. by replacing the 'organizational readiness' with 'readiness' factor and the other two factors have been remained. This study adapts the findings of Chwelos et al. where 'perceived benefits' has not been included in the context of determinants. This paper classifies EDI influencing factors as 'readiness' and 'external influences'. The 'readiness' determinant comprises trading partners' readiness as well as firm's financial resources, IT sophistication, top management support and personnel

acceptance. Meanwhile, 'external influences' determinant includes competitive pressure, industry pressure, imposition by trading partners and enforcement by government.

2.2 Benefits

Numerous past researches have discussed in details about EDI enormous benefits. In general, EDI implementation could lead to positive impacts into five key business areas i.e. customer service, supply chain relationship, internal production, international business and operational cost [9]. EDI benefits can be classified as operational and strategic benefits [19] where operational benefits give immediate impacts and strategic benefits concern on long-term gains for the firm through indirect benefits. EDI benefits can be also defined as direct benefits and indirect benefits [20]. This paper classifies EDI benefits into five group namely administrative costs, information quality, operations management, strategic advantages and transaction speed [5].

2.3 Extent of Integration

In general, there are two dimensions in EDI integration i.e. internal and external [7] [20]. Internal integration refers to the variety of applications interconnected through EDI while external integration refers to the number of EDI trading partners. Jun et al. [21] identified four key EDI integration dimensions namely volume (extent of documents handled), diversity (number of distinct document types handled), depth (degree of business process electronic consolidation between two or more trading partners) and breadth (extent EDI connections with external organizations). As for this study, EDI integration was assessed based on the extent of manual intervention during data transmission, for instant the high integration level comprises little or no manual intervention and vice versa to poor integration level [22].

2.4 Impacts on Firm Performance

It has been argued that IT can be used to enhance firm provided it brings organizational changes and it pairs with human and management factors [23]. Firms are prepared to adopt new technologies if they are convinced with potential significant impacts on their performance. Previously, Kim et al. [24] assessed the firm performance as the results from using IT based on four market measures namely product development, sales growth, market share and market development. While Vargas et al. [23] argued to insert elements of variations in market share and clients' perceptions towards the firm into traditional measures (i.e. profitability, productivity, market value) of firm performance. However, the present study examines the impact of EDI benefits on overall firm performance without focusing into any sub-aspects.

3. Methodology

In Malaysia, all companies that engage in international trade (import/ export) have been mandated to implement EDI (known as SMK DagangNet) for customs declaration. Yet, there are companies which hire third-party agents for such purpose (known as indirect users) as well as companies which implement EDI directly from their own premises (known as direct users). In this study, only direct users from two major industrial states were surveyed. Direct users were further classified as “mandated users” and “self-initiated users”; mandated users refer to companies that initially implement EDI due to imposition while self-initiated users refer to companies that initially implement EDI based on their own initiatives. As a target population for this study, a total of 1200 manufacturers meet this criterion (i.e. direct users) was drawn from Malaysian Manufacturers Portal and Directory (<http://www.e-directory.com.my/>). Of 1200, only 300 manufacturers were randomly chosen. Mail surveys on the chosen 300 Malaysian manufactures were the main data collection method. All respondents were contacted earlier via telephone to solicit their agreement to answer the questionnaire. Only 85 respondents agreed to participate in this survey. The questionnaire was mailed together with reply-paid envelope. Four rounds of follow-up were executed via e-mail to ensure the substantial return rate. Yet, only 71 completed questionnaires were returned. The collected questionnaires were analyzed by using SPSS (Statistical Package for Social Science) to produce reliable results.

Reliability and validity tests were performed for the data collected. The construct reliability was assessed by computing Cronbach’s alpha where items used for the rest analysis showed an adequate level of reliability with their alphas amounting to 0.89. The content validity was enhanced by using the measures already validated in previous studies and by conducting the pilot test among potential respondents and university’s lecturers. The convergent analysis test was performed using the principal component factor analysis on items that measured EDI determinants and benefits. All items of determinants and benefits had factor loading values greater than 0.6, thus none of items were abandoned from further analysis. This signals all final items, which had factor loading values greater than 0.5, have represented the convergent validity. For determinants, only one factor was generated which explains 92.04% of overall variance. While for benefits, five factors were generated which explain 89.05% of overall variance; the generated factors were renamed as costs and information quality-related benefits; time-based benefits; strategic benefits; external benefits; and operational benefits. These generated factors for EDI benefits are alike typology of Bergeron and Raymond [5].

4. Analysis and Discussions

Of 71 respondents, 53 (74.6%) admitted that they are mandated users where 94% of them implement EDI at poor and average level (Table 1). This result support findings of Leng Ang et al. [12] who also found that there were more mandated users (63%) than self-initiated users (37%).

Table 1: Frequency Analysis

Category	n = 71	100%
Types of EDI users		
Mandated	53	74.6
Self - initiated	18	25.4
Level of EDI integration		
High	4	5.63
Moderate	25	35.21
Poor	42	59.16

Mann-Whitney *U* test was employed to measure the difference between levels of integration by type of EDI users (Table 2). Findings indicate level of EDI integration differs by type of users as mean value of mandated users (46.85) considerably higher than mean of self-initiated users (31.42) as well as with lower value of $p = 0.008$. For instant, mandated users might stall EDI implementation with a single set of transaction and reluctant to further integrate EDI into other applications while the self-initiated users opt to integrate EDI into other applications over time as they have fully realized the efficiency and enormous benefits of EDI [16] [20].

Table 2: Mann-Whitney U Test

Type of EDI Users	n = 71	Mean	p (2-tail)	Z
Self-Initiated	23	31.42	0.008	-6.353
Mandated	48	46.85		

Table 3 shows the results of mean value for EDI determinant. Results indicate that top management support sufficient financial resources, enforcement by government and compatibility with existing system were the most significant factors. These results were supported by many past literatures [2] [7] [10] [11] [12]. Cooperation and support of top management is vital for success EDI implementation as it entails significant changes at organization and in trading partners’ relationship. Further, EDI implementation also requires substantial investment thus insufficient financial resources would inhibit further integration. As discussed earlier, many companies were mandated to implement EDI by government for the sake of customs declarations, thus without willingness these companies implement EDI.

Table 3: Mean of EDI Determinants

Determinants	Mean
Top management support	4.52
Sufficient financial resources	4.49
Enforcement by government	4.15
Compatibility with existing system	4.03
Availability of technical persons	3.97
Imposition by trading partners	3.49
Influences from industry players	3.15
EDI training and education for users	3.06
Staff acceptance	2.78
Assistances from EDI trading partners	2.43
Promotion from EDI vendors	1.65

Furthering the analysis, mean value for EDI benefits was computed based on the generated factors from factor analysis (Table 4). Results indicate respondents receive more time-based benefits, followed by benefits-related to reductions in costs and information quality. Respondents claimed to receive least strategic and operational benefits. Strategic and operational benefits are the enhanced EDI benefits which can be fully enjoyed when there is high level of integration [5] [9]. Such benefits were least received as the extent of EDI integration among surveyed respondents is poor and average (94%, Table 1). Benefits-related to reductions in costs and information quality as well as time-based benefits are the direct or minimal benefits [9] [20] which can be enjoyed with average level of EDI integration as per surveyed respondents.

Table 4: Mean of EDI Benefits

Benefits	Mean
Costs and information quality	3.62
Time-based	3.75
Strategic	2.83
External	3.32
Operational	2.70

Accordingly, one way analysis of variance (ANOVA) was conducted to further determine differences in received benefits based on level of integration. Results indicate ($F=5.290$, $p=0.007$) that received benefits differ by level of integration. This result confirms findings of several studies which have demonstrated attainment of EDI benefits subject to the extent of EDI integration [5] [6] [7].

Mean value was computed for the impacts of EDI benefits on overall firm performance. Impacts of EDI benefits on firm performance had mean value of 3.32 with standard deviation of 1.24. This indicates, in average, surveyed respondents claimed EDI benefits have moderate positive impacts on overall firm performance. In order to further examine the relationship between EDI benefits and overall firm performance that results from EDI implementation, Pearson correlation was conducted. Coefficient value of 0.317 indicates there is a significant relationship between EDI benefits and firm performance at significant level of 0.01.

In order to measure the strength of relationship between five groups of EDI benefits and firm performance, a stepwise regression analysis was employed (Table 5). Of five groups of benefits, operational benefits were excluded. Values of $F=21.899$, which is not at all close to unity, and lower $p=0.000$, indicate that EDI benefits had significant impacts on firm performance. The highest beta coefficient of 0.598 among the four groups of benefits indicates that time-based benefits are the most important benefits that yield significant impacts on overall firm performance. Further, 54.4% of the variance is shared by four groups of benefits indicating that model is effective in predicting overall firm performance.

Table 5: Stepwise Regression Analysis

Benefits	Beta	t	p	F	Adj. R ²
Costs and information quality	0.424	4.602	0.000	21.889	0.544
Time-based	0.598	7.281			
Strategic	-0.189	-2.231			
External	-0.181	-2.019			

This result further validates the findings of several studies which have also demonstrated impacts of EDI benefits on firm performance [8] [17] [18]. EDI benefits have been proven to yield significant impacts on firm performance in the context of five major functional areas including accounting, material resource planning, order entry, production scheduling, purchasing and shipping [22].

5. Conclusions

This study was based on a survey on 71 manufacturers. Present level of EDI integration, key EDI determinants and key EDI benefits were identified. Further, the relationship between EDI benefits and overall firm performance was also examined. Results indicate level of EDI integration is merely poor and average (94%). It is something to be concerned on as companies would fail to enjoy 70% of potential EDI benefits if EDI is not integrated [10]. Thus, without enhanced benefits, EDI would probably not yield significant impacts on firm performance. For determinants, top management support, sufficient financial resources and enforcement by government were identified as most influencing factors. While, time-based benefits were identified as the most significant benefits for the surveyed respondents followed by benefits related reductions in costs and information quality.

Clearly, the findings of this study cannot be generalized as this study was conducted merely on manufacturers that located at two particular states instead of all users from different economic sectors throughout Malaysia. Further, the sample size was also not adequate to generalize the findings. Mail surveys have also limited the face-to-face contact with respondents and thus the underlying responses of respondents have failed to identify. Yet, this study is expected to be a significant contribution for local

EDI/ IT literatures as well as a starting point for further investigations.

It is recommendable to examine the factors that inhibit EDI implementation. Future researches can be conducted across different economic sectors and nationwide. A future research shall examine EDI implementation by firm's size i.e. small and medium-sized enterprises (SMEs) and large and multinational companies (LMNs). Further, both quantitative and qualitative methods are suggested to employ to produce comprehensive findings. A longitudinal study should be employed to track the changes in EDI implementation process as well as users' perceptions and acceptance towards EDI over time.

References

- [1] McNurlin, B.C. (1987). The Rise of Co-operative Systems. *EDP Analyzer*. 25(6): 1-16.
- [2] Ngai, E.W.T. and Gunasekaran, A. (2004). Implementation of EDI in Hong Kong: An Empirical Analysis. *Industrial Management and Data Systems*. 104(1): 88-100.
- [3] Chwelos, P., Benbasat, I. and Dexter, A.S. (2001). Empirical Test of an EDI Adoption Model. *Information Systems Research*. 12(3):304-321.
- [4] Swatman, P.M.C and Swatman, P.A (1991). Electronic Data Interchange: Organizational Opportunity, Not Technical Problem. "DBIS' 91" 2nd Australian Conference on Database and Information Systems. February 2- 3. University of New South Wales.
- [5] Bergeron, F. and Raymond, L. (1997). Managing EDI for Corporate Advantage: A Longitudinal Study. *Information and Management*. 31: 319-333.
- [6] Premkumar, G., Ramamurthy, K. and Nilakanta, S.(1994). Implementation of Electronic Data Interchange: An Innovation Diffusion Perspective. *Journal of Management Information Systems*. 1(2): 157-186.
- [7] Iacovou, C.L., Benbasat, I. and Dexter, A.S. (1995). Electronic Data Interchange and small organizations. Adoption and impact of technology. *MIS Quarterly*. 19(4): 465-485.
- [8] Vega, L., Salvador, R. and Guerra, L. (1996). *Determinants and Outcomes of Electronic Data Interchange Integration*.
- [9] Emmelhainz, M.A. (1994). Electronic Data Interchange in Logistics. In: Robeson, F.J. and Copacino C.W. (1994). *The Logistics Handbook*. New York: The Free Press. 737-766.
- [10] Angeles, R., Ravinder Nath and Hendon, W.D. (1998). An Empirical Investigation of the Level Electronic Data Interchange (EDI) Implementation and its Ability to Predict EDI System Success Measures and EDI Implementation Factors. *International Journal of Physical Distribution and Logistics*. 28(9): 773-793.
- [11] Gengeswari, K. and Abu Bakar, A.H. (2006). *Electronic Data Interchange (EDI): Views from Malaysian Manufacturing Companies*. Business IT Conference. 21-23 June. University Institute Technology MARA, Kuala Lumpur, Malaysia: Upcoming
- [12] Leng Ang, C., Razman, M.T. and Rusdi, M. (2003). An Empirical Study on Electronic Commerce Diffusion in the Malaysian Shipping Industry. *Electronic Journal on Information Systems in Developing Countries*. 14(1): 1-9.
- [13] Business Times. *Customs Electronic Clearing System Set to Go Nationwide by December*. 26/11/03.
- [14] Noor Apandi Osnin (2003). *EDI in Transportation*.
- [15] Mahfuzah Kamsah and Wood-Harper, A.T. (1997). A Multi-theoretical Framework for the Study of EDI Diffusion in Malaysia. Management of Technology for Industrialization- Malaysian Science and Technology Congress' 97. 16-17 October. Johor, Malaysia.
- [16] Chau, Y.K. (2001). Inhibitors to EDI Adoption in Small Businesses: An Empirical Investigation. *Journal of Electronic Commerce Research*. 2(2): 78-88.
- [17] Lee, S. and Lim, G. G. (2003). The Impact of Partnership Attributes on EDI Implementation Success. *Information and Management*. 42: 503-516
- [18] Van Heck, E. and Ribbers, M.P. (1999). The Adoption and Impact of EDI in Dutch SMEs. Proceedings of the 32nd Hawaii International Conference on System Sciences.
- [19] Mukhopadhyay, T. and Kekre, S. (2002). Strategic and operational benefits of electronic integration in B2B procurement processes. *Management Science*. 48(10). 1301- 1313.
- [20] Elbaz, D. (1998). *Electronic Data Interchange: An Assessment of the Factors Leading to EDI Adoption*. Concordia University: Master Thesis.
- [21] Jun, M., Cai, M. and Peterson, R.T. (2000) EDI Use and Participation Models: from the Interorganizational Relationship Perspective. *Industrial Management & Data Systems*. 100 (9): 412 – 420.

- [22] Morell, J.A., Neal, W. and Fries, V. (1995). *Promoting Electronic Data Interchange: Building a Foundation for Support to Small Businesses*. Industrial Technology Institute.
- [23] Vargas, A., Hernandez, J.M. and Bruque, S. (2003). Determinants of Information Technology Competitive Value. Evidence from a Western European Industry. *Journal of High Technology Management Research*. 14: 245-268.
- [24] Kim, D., Cavusgil, T.S. and Calantone, J.R. (2005). The Role of Information Technology in Supply-chain Relationships: Does Partner Criticality Matter? *Journal of Business and Industrial Marketing*. 20(4/5): 169-178.

Openness and Productivity Growth in Malaysia

Siti Sumiati¹ and Assoc Prof. DR Azmi Abd. Rahman²

Faculty of Management and Human Resources Development
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Telp: +60-7-5537447, E-mail: siti_sumiati@yahoo.com

Abstract

A major empirical interest in growth studies is whether permanent changes in economic fundamentals affect the long-run growth rate. This paper examines the role of international trade and investment in the process of economic growth and focusing on the contribution of openness variable to productivity growth. Openness variable means open to international trade and investment, thus it involves the effect of the change of export, import and Foreign Direct Investment (FDI). The Dickey-Fuller (DF) and the Augmented Dickey Fuller (ADF) statistics is applied to test the unit roots and to identify the stochastic trends, An Error Correction Model is applied to analyze the short run and the long run dynamics of the model with the information from the cointegrating relationship. The estimation result found that there is a long run relationship between total factor productivity and openness variables.

Keywords: Total Factor Productivity, Openness, unit root test, cointegration, Error Correction Model.

1. Introduction

Productivity growth is a key indicator for long run economic growth. From the economic growth theory we know that the accumulation of the inputs such as labour and capital causes the economy growing to a new steady state in the transition process. However to maintain an economic growth in the long run there must be a technological progress. Since Malaysia is forward oriented, it has argued that there was a positive relationship between openness and economic growth.

Malaysia, as one of the fastest developing economies in this region, enjoyed rapid growth as reflected in GDP, which grew at an average of 6.7% during 1971-90 (Outline Perspective Plan [OPP1]) and 7.0% in 1991-2000 (OPP2). The economy grew at an average growth of 7.0% per annum and achieved the target set in OPP2, although the economy was affected by the Asian financial crisis in 1997-98 which saw GDP register negative growth of 7.5%. In 2001, the Malaysian economy registered a lower growth of 0.4% due to the slowdown of the world economy as compared to growth of 8.5% in 2000. Under OPP3, it is envisaged that GDP will grow at an average 7.5% per annum. The years after the economic crisis saw better economic performance due to demand from domestic consumption emerging from the recovery in private-sector activities and the impact of government measures to consolidate for future growth. In addition, the manufacturing sector continued to maintain its position as the leading sector, contributing 31.5% of GDP in 2001.

In the past, Malaysian economic growth was stimulated by investment, with capital accumulation contributing more than 50% to productivity growth (Abd.Wahab, 2004). The Seventh Malaysia Plan (1996-2000), Eighth Malaysia Plan (2001-05) and Ninth

Malaysia Plan (2006-2010) stipulated a shift in national development strategy from one that is input driven to productivity driven. This strategy is further emphasized in OPP3 (2001-10), which focuses on achieving sustainable growth through productivity improvement with the emphasis on increasing the contribution of total factor productivity (TFP) to output. Table 1 depicts the contributions to economic growth by labour, capital, and TFP for different OPP periods. While Figure 1 show the contribution of TFP growth to GDP growth (%) from 1980 until 2001.

Table 1. Contribution of factors of production, TFP, and GDP growth to economic growth (%).

Period	labour	Capital	TFP	GDP growth
OPP1 (1971-90)				
% Contribution	2.4	3.4	0.9	6.7
%of Total	36.1	50.9	13.0	100.0
OPP2 (1991-2000)				
% Contribution	1.7	3.5	1.8	7.0
%of Total	24.3	50.2	25.5	100.0
OPP3 (2001-2010)				
% Contribution	1.6	2.7	3.2	7.5
%of Total	20.9	36.6	42.5	100.0

Source: Third Outline Perspective Plan (2001-2010)

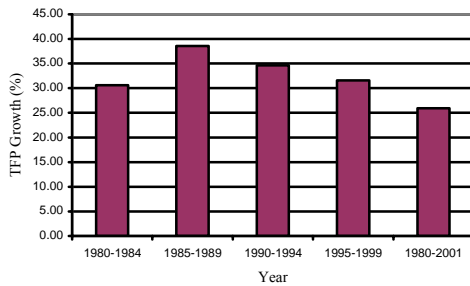


Figure 1. Contribution of TFP growth to GDP growth (%)
Source: APO (2004)

Malaysia is an important trading partner for the United States. In 1999, two-way bilateral trade between the U.S. and Malaysia totaled U.S. \$30.5 billion, with U.S. exports to Malaysia totaling U.S.\$9.1 billion and U.S. imports from Malaysia increasing to U.S.\$21.4 billion. Malaysia was the United States' 12th-largest trading partner and its 17th-largest export market. During the first half of 2000, U.S. exports totaled U.S.\$5 billion, while U.S. imports from Malaysia reached U.S.\$11.6 billion.

The Malaysian Government encourages Foreign Direct Investment (FDI). According to Malaysian statistics, in 1999, the U.S. ranked first among all countries in approved FDI in Malaysia's manufacturing sector with approved new manufacturing investments totalling RM5.2 billion (US\$1.37 billion). Principal U.S. investment approved by the Malaysian Investment Development Authority (MIDA) was concentrated in the chemicals, electronics, and electrical sectors. The cumulative value of U.S. private investment in Malaysia exceeded \$10 billion, 60% of which is in the oil and gas and petrochemical sectors with the rest in manufacturing, especially semiconductors and other electronic products.

Based on the background and the importance of the openness variable in the Malaysia's economic growth, this paper will examine the relationship between openness and total factor productivity growth from 1980 – 2000.

2. Openness and Economic Growth

From Smith (1776) to Ohlin (1965), classical and neoclassical economists tried to explain the gains from trade and how they influence the nation's welfare. However, it is not until the late 20th century that the development of a variety of growth theories helps economists to focus on the channel through which international trade or openness influences economic growth. With the development (during the 1980s) of endogenizing long run growth models (such as Romer, 1986; 1990), Lucas (1988), Prescott and Boyd (1987) economists now have more formal techniques with which to explore the relationship between international openness and economic growth.

There have been many theories that work as well as empirical studies focusing on the topics of openness and economic growth. The channels, which suggested that the openness contributes to economic growth, can be

summarized as shown in Figure 2.1. There are three channels where international openness may affect the economic growth: the terms of trade effect, the trade effect, and the foreign direct investment (FDI) effect. These effects may involve a short run, medium run or long run contribution to economic growth. However, there is also considerable controversy over the empirical results.

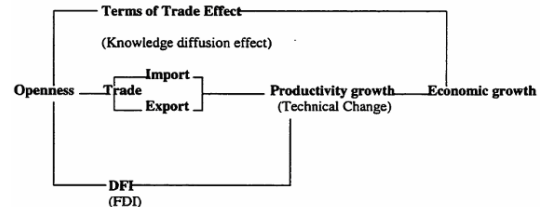


Figure 1. How openness affects economic growth(Source Wang, 2001)

Many economists (World Bank, 1997; Evenson and Westphal, 1999) believe that export-led development strategies improve technical efficiency. One reason is that if buyers want low-cost, better quality products from their suppliers, they may transmit knowledge from other countries' suppliers; thus through participating in export markets, firms may use international best practices and help their productivity growth. The second reason is that, like Schumpeter's idea, to maintain their rents and compete with other international rivals, firms need to lower their cost and hence simulate technology innovation. The third reason is that through exports firms are exposed to the international best practice environment, hence fostering the learning by doing effect.

There are also some other two variable Granger causality tests analyses that obtained ambiguous results, such as Jung and Marshall (1985), Hsiao (1987) and Dodaro (1993), while Chow (1987) confirmed causality between exports and output growth for Asian Tigers (see Van den Berg (1996) for a more detailed description). Bernard and Jensen (1999) have shown that for U.S. manufacturing sector the overall contribution of exporters to manufacturing productivity growth far exceeds their shares of employment and output.

3. Measurement of Total Factor Productivity

The measurement of TFP is related to the specification and estimation of the production function. The model used here attempts to estimate TFP based on the growth accounting framework pioneered by Solow (1957). TFP is the residual and is considered to represent the rate of technical change. Growth accounting is a statistical technique for dividing the growth of output into two sources: the growth of a variety of inputs, such as investment in physical capital or increases in size of labor; and the growth of TFP on the other. TFP is the economic and technical efficiency with which resources such as capital and labor are converted into output.

Based on this framework, the mathematical presentation of the model used is given below. The equations used to compute the sources of economic and

productivity growth are derived using a production function:

$$Q = f(K, L) \tag{Eq. 1}$$

where Q = output or GDP, K = capital, and L = number of workers.

By including a time variable (t , assumed due to technical progress), the resulting shifts of the production function can be represented by:

$$Q_t = f(K_t, L_t, t) \tag{Eq. 2}$$

thus implying that the same input quantities yield a different outputs at different points of time.

Assuming that technical progress is both neutral and disembodied (Solow, 1957), the production function in Eq. 2 can be expressed as:

$$Q_t = A(t) \cdot f(K_t, L_t) \tag{Eq. 3}$$

where Q_t , K_t , and L_t = output and factor inputs during period t , and $A(t)$ = technical progress or TFP as a function of time.

Differentiating Eq. 3 with respect to time and denoting the derivatives by putting a dot over the variable,

$dQ/dt = \dot{Q}$ we have

$$\dot{Q} = \dot{A} \cdot f(K_t, L_t) + A \cdot \frac{\partial f}{\partial K} \cdot \dot{K} + A \cdot \frac{\partial f}{\partial L} \cdot \dot{L} \tag{Eq. 4}$$

Dividing throughout by Q leads to an expression for the proportionate rate of change in output:

$$\frac{\dot{Q}}{Q} = \frac{\dot{A} \cdot f(K_t, L_t)}{Q} + A \cdot \frac{\partial f}{\partial K} \cdot \frac{\dot{K}}{Q} + A \cdot \frac{\partial f}{\partial L} \cdot \frac{\dot{L}}{Q} \tag{Eq. 5}$$

$$\frac{\dot{Q}}{Q} = \frac{\dot{A} \cdot f(K_t, L_t)}{Q} + A \cdot \frac{\partial f}{\partial K} \cdot \frac{K}{Q} \cdot \frac{\dot{K}}{K} + A \cdot \frac{\partial f}{\partial L} \cdot \frac{L}{Q} \cdot \frac{\dot{L}}{L} \tag{Eq. 6}$$

Solow (1957) assumed that factor inputs are paid the value of their marginal products under competitive equilibrium conditions, so that:

$$\frac{\partial Q}{\partial K} = A \cdot \frac{\partial f}{\partial K} = \frac{r}{p}$$

$$\frac{\partial Q}{\partial L} = A \cdot \frac{\partial f}{\partial L} = \frac{w}{p} \tag{Eq. 7}$$

where p = prices of output, r = prices of capital inputs, and w = prices of labor inputs.

$$\frac{\dot{Q}}{Q} = \frac{\dot{A}}{A} + \frac{rK}{pQ} \cdot \frac{\dot{K}}{K} + \frac{wL}{pQ} \cdot \frac{\dot{L}}{L} \tag{Eq. 8}$$

In Solow's notation, the shares of capital and labor are denoted by $wK = r.K/p.Q$ and $wL = W.L/p.Q$, respectively, and thus with this assumption Eq. 8 becomes:

$$\frac{\dot{Q}}{Q} = \frac{\dot{A}}{A} + w_K \cdot \frac{\dot{K}}{K} + w_L \cdot \frac{\dot{L}}{L} \tag{Eq. 9}$$

Further, assuming constant returns to scale, where a percentage change in inputs leads to the same percentage change in output, the following holds: $WK + WL = 1$ Therefore Eq. 9 becomes:

$$\frac{\dot{Q}}{Q} = \frac{\dot{A}}{A} + w_K \cdot \frac{\dot{K}}{K} + (1 - w_K) \cdot \frac{\dot{L}}{L} \tag{Eq. 10}$$

Where:

\dot{Q} / Q = proportionate rate of change in output,

\dot{A} / A = proportionate rate of change in technical progress or TFP,

\dot{K} / K = proportionate rate of change in capital, and

\dot{L} / L = proportionate rate of change in labor.

4. Data Sources

Openness is measured by using its proxy of exports, imports, and FDI. Both export and import data are drawn from Department of Statistic Malaysia. The data for exports and imports for each industry within the manufacturing sector was drawn from the Monthly Statistics of Exports and Imports, Malaysia. Foreign direct investment (FDI) data for the whole economy is drawn from the Malaysian Industrial Development Authority (MIDA) Malaysia.

In this study, human capital is defined as the number of workers classified at the managerial and professional and technical and supervisory (e.g. technicians) level divided by numbers of persons engaged during December or the last pay period of the reference year, collected from Department of Statistics Malaysia in Annual Survey of Manufacturing Industries.

The other data is the relative ratio of U.S. per capita GDP to Malaysia per capita GDP as a proxy of the technical gap between the leading technology country and Malaysia. These data sets were mainly drawn from International Financial Statistics and the full sample period is 1980 to 2000.

5. Estimation Technique of Error Correction Models (ECM)

Most of the earlier for the determinants of economic growth and productivity growth relied upon standard econometric techniques, such as using the classical ordinary least square (OLS) as an estimation technique without investigating the presence of a stochastic trend of nonstationary data. The deficiency and failure of those techniques have been known as "spurious regression", which is related to the nonstationary behavior of economic time-series data. The t -statistic estimated by the OLS regression of nonstationary data does not follow the standard distribution, leading to spurious regression (Enders, 1995; Johnston & Dinardo, 1997).

The presence of a stochastic trend is determinate by unit root test in the polynomials of lag operators of the time series. The test statistic suggested by Dickey and Fuller (1979) serve to detect stochastic trends in time-series data. Thus, since the regression analysis in this study involves time-series data, we must test the presence of any stochastic trend. If a stochastic trend exists, detrended values of the time series with appropriate different data

must be used so as to make the regression analysis meaningful.

Therefore, in this study, the Dickey-Fuller (DF) and the augmented Dickey-Fuller (ADF) statistics are applied to test the unit roots and to identify the stochastic trends. In addition, to avoid the number of lags of variables chosen arbitrarily, the Enders (1995) method of determining the appropriate lag length in the regression equation is employed for this analysis. Moreover, the error-correction model (ECM), suggested by Engle and Granger (1987) is applied to identify the nature of the determinants of the long-run productivity growth model in Malaysia.

This section deals with the analysis of the presence of stochastic trend of nonstationary data based on time-series properties of these data in the natural logarithmic form (Ln). The Dickey-Fuller (DF for the appropriate lag length, $p=0$) and the augmented Dickey-Fuller (ADF) unit root tests for the random walk drift and with/without linear time trend of these data are used in the following equation:

$$\Delta \ln(x)_t = \mu + \alpha t + \beta \ln(x)_{t-1} + e_t \quad (\text{Eq. 11})$$

$$\Delta \ln(x)_t = \mu + \beta \ln(x)_{t-1} + e_t \quad (\text{Eq. 12})$$

$$\Delta \ln(x)_t = \mu + \alpha t + \beta \ln(x)_{t-1} + \sum_{i=1}^p \eta_i \Delta \ln(x)_{t-i} + e_t \quad (\text{Eq. 13})$$

$$\Delta \ln(x)_t = \mu + \beta \ln(x)_{t-1} + \sum_{i=1}^p \eta_i \Delta \ln(x)_{t-i} + e_t \quad (\text{Eq. 14})$$

Where: Δ denotes the first-difference operator, x_t is the relevant variable under investigation [HC], [EXP], [IMP], [FDI] and [MA], p is the appropriate lag length, t is the deterministic time trend, and e_t is the white-noise residual. Two important assumptions for the estimation of equations (Eq.11) – (Eq.12) are: (i) the logarithmic transformation is correct, and (ii) there are no structural breaks in trend function. To choose the appropriate lag length, due to small sample time-period data in this study, we initially set $p=2$ and then follow the Enders (1995) method. If the t -statistic of lag p from the regression is insignificant at some specified critical value, the regression is re-estimated by using lag length of $p-1$ until the lag length is significantly different from zero.

Next, test the null hypothesis, $H_0: \beta = 0$ versus $H_a: \beta \neq 0$, by comparing the test statistic with the Dickey-Fuller critical value. If H_0 is accepted, x_t is non-stationary, and thus x_t contains unit root [$x_t \sim I(1)$ process].

In addition, if two roots are suspected, the Dickey-Fuller test based on equation (Eq.11) and (Eq.12) is applied to estimate the following equations:

$$\Delta^2 \ln(x)_t = \mu + \alpha t + \beta_1 \ln(x)_{t-1} + e_t \quad (\text{Eq. 15})$$

$$\Delta^2 \ln(x)_t = \mu + \beta_1 \Delta \ln(x)_{t-1} + e_t \quad (\text{Eq. 16})$$

Where Δ^2 is denotes the second-difference operator. The null hypothesis, $H_0: \beta = 0$ versus $H_a: \beta \neq 0$, is tested by comparing the test statistic with the Dickey-Fuller critical

values (Enders, 1995). If H_0 is accepted, thus $x_t \sim I(1)$ process. If H_0 is rejected, we go on to determine whether there is a single-unit root by estimating the following equations:

$$\Delta^2 \ln(x)_t = \mu + \alpha t + \beta_1 \Delta \ln(x)_{t-1} + \beta_2 \Delta \ln(x)_{t-2} + e_t \quad (\text{Eq. 17})$$

$$\Delta^2 \ln(x)_t = \mu + \beta_1 \Delta \ln(x)_{t-1} + \beta_2 \Delta \ln(x)_{t-2} + e_t \quad (\text{Eq. 18})$$

If the results show that the parameters β_1 and/or β_2 are significantly different from zero, then there are no two-unit roots. Thus, under the null hypothesis of a single-unit root, $\beta_1 < 0$ and $\beta_2 = 0$.

Next, the cointegration test is applied to examine the long-run relationship between $\ln(TFPG)$ and $\ln(HC)$, $\ln(EXP)$, $\ln(IMP)$, $\ln(FDI)$, and $\ln(MA)$ that are integrated to the same order $I(1)$ for the total factor productivity growth based on the following equation:

$$\Delta \hat{e}_t = \gamma \hat{e}_{t-1} + v_t \quad (\text{Eq. 19})$$

Where $\hat{e}_t = [\ln(TFPG)_t - \sum_{i=1}^r \Phi_i' \ln(Z_i/L)_t]$, or is the estimated residual from the labour productivity growth regression equation: Φ_i is the coefficients of the cointegrating vector; r is the number of explanatory variables in the cointegrating vector; $'$ denotes the transpose operator; Z_i denotes HC , EXP , IMP , FDI or MA ; and v_t is the white-noise residual. However, if v_t is not the white noise, the equation (Eq.19) is replaced by the following autoregressive processes. Then the augmented Dickey-Fuller (ADF) test is applied in the following equation:

$$\Delta \hat{e}_t = \gamma \hat{e}_{t-1} + \sum_{i=1}^p \Psi_i \Delta \hat{e}_{t-i} + v_t \quad (\text{Eq. 20})$$

To test the null hypothesis, $H_0: \gamma = 0$ versus $H_a: \gamma \neq 0$, in equation (Eq.19) and (Eq.20), the t -statistic is compared with the MacKinnon (1991) critical values of no cointegration, applied for several observations. The statistically significant negative t -statistic leads to reject H_0 , and then the residual \hat{e}_t is stationary. In case a group of variables is $I(1)$ process, $\ln(TFPG)$ and $\ln(Z_i/L)$ in the total factor productivity growth equation are cointegrated $CI(1.1)$.

Next, an error-correction model (ECM) is applied to analyze the short-run and the long-run dynamics of the model with the information from the cointegrating relationship. In this analysis, when the variables are cointegrated, the estimated cointegrating vector represents the long-run equilibrium relationship and must be included in the model in order to remove the omitted variable bias. The idea of ECM is that a proportion of the disequilibrium from one period is corrected in the next period. Thus, if $\ln(TFPG)_t$ and $\sum_{i=1}^r \ln(Z_i/L)$ are found to be significantly cointegrated $CI(1.1)$ by the DF or the ADF methods, Engle and Granger (1987) suggest there exists an analogous error-correction model (ECM) in the following form:

$$\Delta \ln(Y_t) = \beta_0 + \sum_{j=0}^r \beta_j \Delta \ln(Z_j/L) + \sum_{j=1}^r \Omega_j \Delta \ln(Y_{t-j}) + \delta [\ln(Y_{t-1}) - \sum \Phi_i \ln(Z_i/L)] + \varepsilon_t \tag{Eq.21}$$

Where Δ denotes the first-difference operator; β_{ij} are the estimated short-run elasticity coefficients; r is the number of explanatory variables in the cointegrating relationship; J is the number of lag length; δ is the estimated coefficient of the lagged error-correction term; Φ_i is the coefficient of cointegrating vector; Ω_j is the coefficient of lagged dependent variable; ' denotes the transpose operator; Z_i represents *HC, EXP, IMP, FDI, or MA* and ε_t is the white-noise residual.

In economic analysis, $\Delta \ln(TFPG)_t$ and $\Delta \ln(Z_i/L)_t$ refer to the total factor productivity growth and the growth of factor Z_i , respectively. Moreover, in this analysis, we initially set the number of lag length, $j=0$ for $\Delta \ln(Z_i/L)$, and $j=1$ for its lagged dependent variable. If the residual ε_t is serially corrected, we re-estimate by increasing the lag length.

The error-correction term is generated from the cointegrating residual. Thus, Δ (speed of adjustment) measures how changes in those variables respond to departures from the long-run equilibrium. The variables $\ln(TFPG)$ and $\sum^r \ln(Z_i/L)$ in the total factor productivity growth equation are cointegrated when the coefficient δ is significantly different from zero. If $\delta \neq 0$, $\sum^r \ln(Z_i/L)$ will lead $\ln(TFPG)$ to approach the long-run equilibrium (growth rate of $[TFPG]_t$, or the total factor productivity growth at time t will respond to the deviation from the long-run equilibrium at time $t-1$. Likewise, if β_{ij} is significantly different from zero, the movements in $\sum^r \ln(Z_i/L)$ will lead the movements in $\ln(Z_i/L)$ in the short run.

Since all of the first-difference variables and the error correction term are stationary ($\Delta \ln[TFPG]_t$ and its lags, $\sum^r \ln(Z_i/L)_t$ and its lags, and $\hat{\varepsilon}_{t-1}$ are $I(0)$ process), the ordinary least square (OLS) method will be appropriate and provide unbiased estimates of standard errors. Thus, in this study we estimate all parameters by using the OLS method. Also, the cointegrating relationship provides the t -statistic and F -statistic with a normal distribution, and thus we can analyze the regression without misleading and spurious results (Gujarati, 2003).

In this analysis, the EViews 5.0 econometric software program is employed to provide the results for the estimation techniques of Dickey-Fuller (DF) and the augmented Dickey-Fuller (ADF) unit root tests, the cointegration test, and the OLS method in the error-correction model (ECM).

The coefficients of interest are β_{ij} (the short-run elasticity) and Φ_i (the long-run elasticity), which are hypothesized to be significantly positive.

6. The Econometric Results

Test of DF, ADF, and Phillip Perron to detect the presence of unit roots are shown in table 2, table 3 and table 4. The presence or absence of cointegrating relationships in the set data also tested and the result is shown in table 5.

Table 2. Test of Dickey-Fuller and Augmented Dickey-Fuller without trend

Variables	D-F Test		
	level	1 st Difference	2 nd Difference
LnTFPG1	-3.3469*	-5.3145*	-6.2953*
LnEXP	-0.3553	-3.2295*	-5.368*
LnIMP	0.0336	-2.9580*	-3.8344*
LnFDI	-1.2388	-2.9617*	-5.2653*
LnHC	0.2350	-3.7075*	-3.8613*
LnMA	-2.2086*	-3.2612*	-4.9431*

Variables	ADF Test		
	level	1 st Difference	2 nd Difference
LnTFPG1	-2.7323	-5.3025*	-5.3473*
LnEXP	1.4221	-4.1882*	-5.6378*
LnIMP	0.0070	-2.9488*	-3.7072*
LnFDI	-1.2575	-2.9238*	-5.2858*
LnHC	-0.0703	-3.9932*	-4.2530*
LnMA	-2.2226	-3.2362*	-4.2970*

Notes: * means significant at 5% level
** means significant at 10% level

Table 3. Test of Dickey-Fuller and Augmented Dickey-Fuller with trend

Variables	D-F Test		
	level	1 st Difference	2 nd Difference
LnTFPG1	-4.2497	-3.9519*	-7.5514*
LnEXP	-2.4810	-4.7246*	-4.8228*
LnIMP	-2.9422	-3.9870*	-4.8225*
LnFDI	-2.3026	-3.9835*	-5.4283**
LnHC	-3.7747	-2.3744*	-4.1143*
LnMA	-2.2446	-3.2892*	-4.3352*

Variables	ADF Test		
	level	1 st Difference	2 nd Difference
LnTFPG1	-3.3010	-3.9519*	-6.5963*
LnEXP	-2.4896	-4.7415*	-4.2736*
LnIMP	-3.1204	-2.9842*	-4.2491*
LnFDI	-2.2452	-2.9341*	-3.3327*
LnHC	-3.5695	-3.8175*	-4.0601*
LnMA	-2.1857	-3.9384*	-4.1163*

Table 4: Phillip Perron Test Statistic

Variables	Phillip Perron Test		
	level	1 st difference	2 nd difference
LnTFPG1	-3.3368*	-5.7123*	-124.4869*
LnEXP	1.4221	-3.9952*	-5.3590*
LnIMP	-0.1722	-3.2612*	-3.7072*
LnFDI	-1.4641	-2.923793*	-5.285829*
LnHC	-0.4198	-7.8621*	-27.1059*
LnMA	-2.4988	-3.2362*	-5.9973*

In order to test for presence of stochastic non-stationary in the data, the integration order of individual time series, using the DF, ADF and PP test for the presence of unit roots (Table 2, 3 and 4 respectively). From DF, ADF and PP test indicate that, the null hypothesis of a unit root is not rejected at the 5%

significant level. To verify the order of Integration $I(1)$, the presence of unit roots in the first difference of the variables tested but no unit roots in the first difference series was found. Clearly the results indicate that all variables are stationary in their first-differences.

Table 5. Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.979899	164.4723	95.75366	0.0000
At most 1 *	0.852582	90.23970	69.81889	0.0005
At most 2 *	0.811658	53.86454	47.85613	0.0123
At most 3	0.518248	22.14410	29.79707	0.2906
At most 4	0.315713	8.267912	15.49471	0.4372
At most 5	0.054249	1.059740	3.841466	0.3033

Trace test indicates 3 cointegrating eqn(s) at the 0.05 level

* denotes rejection of the hypothesis at the 0.05 level

**MacKinnon-Haug-Michelis (1999) p-values

As the null hypothesis of unit roots can not be rejected, multivariate models can be build to enable investigation of the presence or absence of cointegrating relationships in the set data (Table 5). The empirical findings indicate the presence of three cointegrating vector. The null hypothesis that the TFPG, export, import, FDI, Human Capital and ratio of U.S. per capita GDP to Malaysia per capita GDP are not cointegrated ($r=0$) against the alternative one or more cointegrating vector ($r>0$) is rejected, since the $\lambda_{\max(0)}$ statistic greater than critical value at 5% significance level. However, the λ_{\max} and λ_{trace} statistic suggest no more than one cointegrating vector, since H_0 of $r \leq 1$ is not rejected, as $\lambda_{\max(1)}$ is less than the critical value at 5% significant level.

We use TFPG as dependent variable in the model with 2 lags. The selection of the model based on minimize of AIC, and the results as follows:

$$\begin{aligned} \Delta\text{TFPG} = & -82.95571 + 1.69163\Phi_{t-1} - 0.577146\Delta\text{TFPG}(-1) \\ & (-50.3182) \quad (0.3152) \quad (0.62535) \\ & [-1.64862] \quad [2.90531] \quad [-0.92292] \\ & -0.51532\Delta\text{TFPG}(-2) + 7.980866\Delta\text{EXP}(-1) \\ & (-0.50582) \quad (-4.02457) \\ & [-1.01877] \quad [1.98304] \\ & -4.026672\Delta\text{EXP}(-2) - 3.281835\Delta\text{IMP}(-1) \\ & (-4.5027) \quad (-3.34253) \\ & [-0.89428] \quad [-0.98184] \\ & -4.281546\Delta\text{IMP}(-2) + 2.364354\Delta\text{FDI}(-1) \\ & (-4.8062) \quad (-1.37639) \\ & [-0.89084] \quad [1.71780] \\ & -1.959051\Delta\text{FDI}(-2) + 5.377925\Delta\text{HHC}(-1) \\ & (-1.19593) \quad (-4.16523) \\ & [-1.63810] \quad [1.29115] \\ & + 6.604518\Delta\text{HHC}(-2) + 4.951755\Delta\text{AMA}(-1) \\ & (-3.269) \quad (-4.97116) \\ & [2.02035] \quad [0.99610] \end{aligned}$$

$$\begin{aligned} & -3.028698\Delta\text{AMA}(-2) \\ & (-2.97096) \\ & [-1.01943] \end{aligned}$$

The result shows that the coefficient of Error Correction Term (Φ_{t-1}) is significant, that means at lag 2 there is long-run relationship between total factor productivity and openness variables.

7. Conclusion

The results of unit root tests include DF, ADF and Phillip Perron tests indicate that all variables are stationary in their first-differences. From cointegration test we know that the empirical findings indicate the presence of three cointegrating vector. After Error Correction Model is used, the empirical findings indicate that there is long-run relationship between Total factor productivity and openness variables.

References:

- [1] Ab. Wahab Muhammad. (2004). *National Reports: Malaysia Part II*. In Report of the Asian Productivity Organization (APO) Survey on Total Factor Productivity 2001/2002. APO, ISBN:92-833-7016-3
- [2] Dodaro, Santo (1993). Exports and Growth: A Reconsideration of Causality. *Journal of Developing Areas*, 27:227-244.
- [3] Chow, Peter C.Y. (1987). Causality between Export growth and industrial development – Empirical evidence from the NIC's. *Journal of Development Economics*. 26:55-63.
- [4] Dickey, D., & Fuller, W. A. (1979). *Distribution of the Estimators For Autoregressive Time Series With a Unit Root*. *Journal of American Statistics Association*. 74:427-431
- [5] Engle, R. F., & Granger, C. W. (1987). *Cointegration and Error-Correction: Representation, Estimation and Testing*. *Econometrica*, 55:251-276.
- [6] Evenson, Robert E. and Larry E. Westphal, Behrman, Jere and T.N. Srinivasan, eds. (1995). *Handbook of development economics*, volume 3A. *Handbooks in Economics*, vol. 9, Amsterdam; New York and Oxford: Elsevier Science, North Holland, pages 2209-2299.
- [7] Enders, W. (1995). *Applied Econometric Time Series*. New York: John Wiley & Sons. Inc.
- [8] Gujarati, D. (2003). *Basic Econometrics*. 4th Edition. New York: McGraw – Hill. Inc.

- [9] Hsiao, M. W. (1987). Test of causality and exogeneity between exports and economic growth: The case of Asia NICs. *Journal of Economic Development*, 12, 143-159.
- [10] Johnston, J., & Dinardo, J. (1997). *Econometric Methods*. 4th Edition New York: McGraw – Hill. Inc.
- [11] World Bank, (1989). *Malaysia: Matching Risks and Rewards in a Mixed Economy Programme*. Washington, D. C.: World Bank.
- [12] Lucas, R. E., (1988). *On the Mechanics of Economic Development*. *Journal of Monetary Economics* 22(1):3-42.
- [13] Prescott and Boyd (1987). Dynamic Coalitions: Engines of Growth. *The American Economic Review* vol 77 no 2.
- [14] Romer, P.M. (1986). *Increasing Returns Endogenous and Long Run Growth*. *Journal of Political Economy*, 94:1002-37
- [15] Romer, P.M. (1990). *Endogenous Technological Change*. *Journal of Political Economy*, 98:71-102
- [16] Van den Berg, Hendrik (1996). Trade as the Engine of Growth in Asia: What the Econometric Evidence Reveals. *Journal of Economic Integration*. 11(4), December 1996; pp. 510-538

Posters



Natural Organic Matter (NOM) Fouling in Surface Water Treatment

A.W. Zularisam¹, A.F. Ismail^{2*}, M.R. Salim³

¹Environmental Engineering Focus Group, FKA, Kolej Universiti Kejuruteraan & Teknologi Malaysia, KUKTEM, Gambang, Pahang, Malaysia

²Membrane Research Unit, University Technology of Malaysia, Skudai Johor, Malaysia

³Faculty of Civil Engineering University Technology of Malaysia, Skudai Johor, Malaysia
Tel: 607-5535592, Email:afauzi@utm.my

Abstract

Membrane application in surface water is governed by fouling issue which restricts its widespread application. This paper discusses NOM and its components as the major membrane foulants, possible fouling mechanisms and current techniques used to identify foulants associated with NOM fouling

Keywords: NOM; Fouling mechanism; Membrane foulant

1. Introduction

Natural organic matter (NOM) reactions with free chlorine have been found to have tendency to form disinfection by product (DBPs) such as Trihalomethanes (THMs), Trihaloacetic (THA) and other halogenated organics. DBPs are carcinogens and direct exposure can lead to cancers, miscarriage and nervous system complications. Superior water quality produce by membrane filtration has made this advanced technology possible to replace the conventional treatment processes, which consist of ozonation-precipitation-coagulation-flocculation-chlorination-gravel filtration [1]. Besides that, membrane filtration process does offer extra advantages over conventional treatment such as small footprint, compact module and capable of handling wide fluctuations in feed quality. Membrane filtration processes involving microfiltration (MF) and ultrafiltration (UF) and nanofiltration (NF) in potable water production have increased rapidly for the past decade. MF and UF are employed to remove microparticles and macromolecules, which generally include inorganic particles, organic colloidal (i.e. microorganism) and dissolved organic matters. Dissolved organic matter is ubiquitous in natural surface water and often reclaimed [2] as important factor for both the reversible and irreversible fouling in water filtration. Fouling is the main obstacle in membrane filtration efficiency as it causes reduction in productivity. Productivity decline can be defined as a decrement in flux with time of operation due to the increment of hydraulic resistance. Productivity decline may also be interpreted as a need for additional energy supply to the filtration system so as to keep the system performance constant. Therefore, a fundamental knowledge on the possible foulants and how they cause fouling are essential before any remediation works is carried out. In the present paper we review the characteristics of NOM fouling during surface water filtration that relate to both reversible and irreversible fouling mechanisms. Discussion are also devoted on the current techniques employed to characterize foulants, fouling control methods, feed properties, membrane characteristics, operational conditions and solution

chemistry that have been reported to be potentially influenced the NOM fouling in surface water filtration.

2. Foulants of Surface Water Filtration

Membrane fouling refers to both reversible and irreversible alteration in membrane properties. Reversible fouling means deposition of retained solutes on the membrane surface that generally exists as a gel cake layer. Irreversible fouling refers to adsorption or pore plugging of solutes in and within the membrane pore matrix. Concentration polarizations are the accumulation of retained materials in the boundary layer above the membrane due to osmotic pressure and hydraulic resistance effect. Increment and variation of hydraulic resistances may come from variety of organic substances, inorganic particles, colloids and microorganisms with different fouling behaviours. The fouling behaviour is significantly found to be influenced by various chemical and physical factors of the foulants. The foulant can be characterized according to their molecular structure, surface charge, molecular size and functional groups. One of the most important identified foulant found in surface water filtration is natural organic matter (NOM). NOM waters are a complex mix of particulate and soluble components of both inorganic and organic origin that vary from one source to others [3]. NOM is a heterogeneous mixture with wide ranges in molecular weight (MW) and functional groups (phenolic, hydroxyl, carbonyl groups and carboxylic acid) and is formed by allochthonous input such as terrestrial, vegetative debris and autochthonous such as algae. Natural organic matters that occur in natural brown water are polyphenolic molecules with MW ranging from 5000 to 50000 Dalton [4]. In particular, NOM can be fractionated into three segments; the hydrophobic fraction (humic substances), hydrophilic

and transphilic. The hydrophobic fraction represented almost 50% of dissolved organic carbon (Figure 1) with larger MW. The hydrophilic fraction composed 25 %- 40% of dissolved organic carbon (DOC) with lower MW (polysaccharides, amino acids, protein and etc.) and operationally defined as non-humic fraction. The transphilic fraction comprised approximately 25% of DOC in natural water but with MW in between hydrophobic and hydrophilic fractions. A major fraction of the NOM arises from humic substances and is reported to represent up to 60% to 70% of TOC in soils and 60% to 90% of DOC in most natural water. Fan et al. [5] reported the major fraction (over 50% of DOC) of NOM is composed of humic substances and are responsible for the natural water's colour. A humic substance is the predominant fraction of NOM and generally is divided into three categories, which are humic acid (HA), fulvic acids (FA) and humin. HA and FA are anionic polyelectrolyte with negatively charged of carboxylic acid (COOH⁻), methoxyl carbonyls (C=O) and phenolic (OH) functional groups. Figure 2 and Figure 3 show both models of humic acid and fulvic acids structures. Humic acid is soluble at higher pH normally 10 while fulvic acid is soluble in water under all pH. Humin is naturally exists in black colour and does not soluble in water at any pH (Table 1). Humic fraction has been identified as the major foulant in membrane water filtration, which controls the rate and extent of fouling [6]. It causes more fouling than any other NOM components due to its adsorptive capacity on the membrane surface [7]. Study done by Mallevalle [8] showed organic matrix formed a structure of fouling layer that served as a glue for inorganic constituents. Similar results were reported by Kaiya et al. [9] in analyzing deposited layer formed on a MF hollow fiber during filtration of Lake Kasumigaura water. NOM deposition has been found as the dominant factor causing flux decline along with manganese constituent. Study by Mo Li and Huang Xia [10] on purification of micro-polluted raw water revealed that fouling on the exterior surface was a combined effect of microorganisms and inorganic matter while on the inner surface was mainly due to the biofouling. They found the organic foulants were of low molecular weight and the inorganic was primarily represented by Ca²⁺ element. Their investigation on membrane permeability recovery showed that alkaline cleaning was effective in removing organic foulants while acidic cleaning was more effective on inorganic scales. It has also been shown that hydrophobic fraction of NOM causes much more fouling than hydrophilic fraction [11].

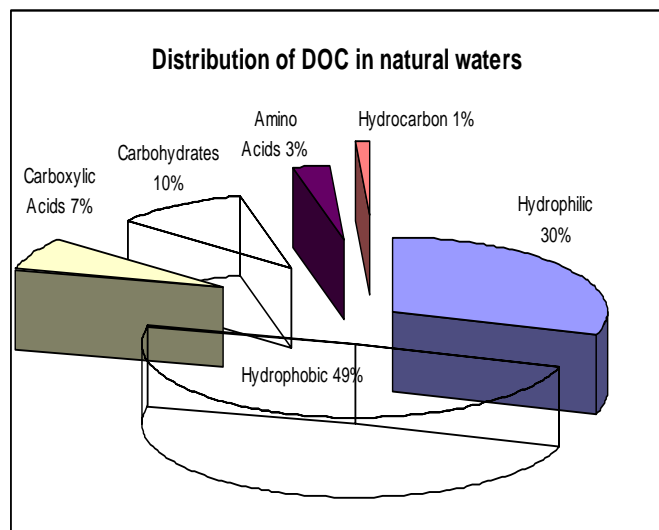


Figure 1. Fraction of NOM in surface water based on DOC [12]

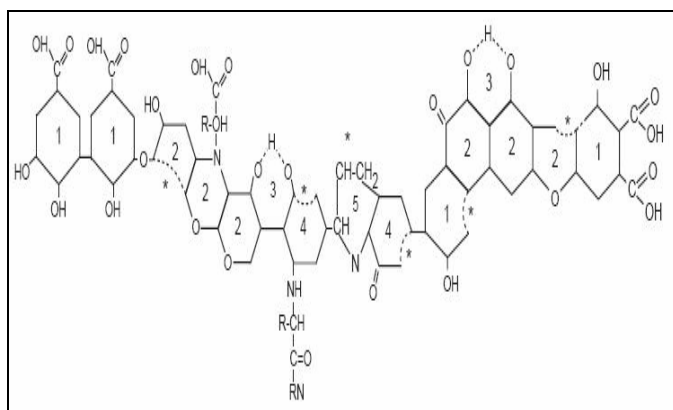


Figure 2. Schematic of humic acid model structure

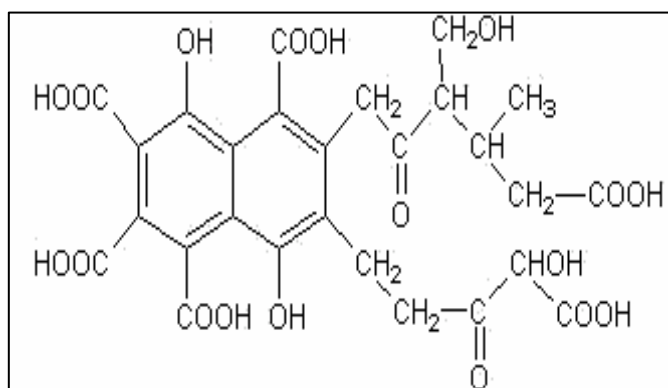


Figure 3. Schematic of fulvic acid model structure

Table 1. Physical and chemical characteristics of humic substances [13]

Fulvic acid		Humic acid		Humic
Light yellow	Yellow brown	Dark brown	Grey black	Black
Increase in degree of polymerization \longrightarrow				
2000	Increase in MW	300,000 and above		
45%	Increase in carbon content	62%		
48%	Decrease in oxygen content	30%		
1400%	Decrease in exchange acidity	500%		
Decrease in degree of solubility \longrightarrow				

They performed NF of a hydrophilic membrane with aquatic NOM using DAX-8 to fractionate the NOM components. Hydrophobic fraction (absorbable to DAX-8) was mainly responsible for the permeate flux decline. On the other hand the hydrophilic component which passed through DAX-8 showed less fouling effect. Humic macromolecules with higher hydrophobicity are found to favourably adsorb onto hydrophobic membrane than hydrophilic fraction [14]. Previous studies done by many researchers show humic substances caused irreversible fouling of membranes [15]. Yuan and Zydney [16] studied humic acid fouling on a 0.16 μm hydrophilic MF and found that aggregate humic acid was responsible for the initial stage of fouling. Furthermore fouling mechanism was substantially due to convective deposition with little internal pore adsorption. This finding is well supported by Schafer et al. [17] who observed humic acid to cause 78% decline in flux compared to fulvic acid (15%). Humic acid was observed to give greater impact on membrane performance (irreversible fouling) than FA and hydrophilic fraction (reversible fouling). This scenario might be due to its high aromaticity properties, adsorptive behaviour, hydrophobic and bigger molecular weight that lead to tendency of fouling. Similar result was also reported by Lahoussine-Turcaud et al. [18] during ultrafiltration observations with several organic and inorganic of Seine River. The flux decline observed was primarily due to humic acid deposition on the membrane surface. The cellulose acetate membrane (hydrophilic) flux was twice times greater than hydrophobic polyethersulfone (PES) during UF of river water. The hydrophilic components were thought to impact water quality less than the humic fraction, however recent studies done by Lin et al. [19] and Carroll et al. [20] have claimed that the non-humic fraction of NOM (hydrophilic and neutrals) materials were responsible in determining the rate and extent of flux decline. Carroll et al. [21] performed MF of hydrophobic hollow fibre membrane with a single water source and concluded that the major cause of fouling was due to the hydrophilic neutral and not the humic substances. Fan et al. [22] reported the order of fouling potential of NOM fraction as hydrophilic neutral > hydrophobic acids > transphilic acids > hydrophilic charged. They found that hydrophobic membrane had the most fouling effect than hydrophilic membrane of similar size suggesting that the fouling mechanism was governed by adsorption. In addition high molecules weight components had been identified as having

the largest impact on membrane fouling compared to smaller DOM. This finding was well supported by study carried out by Lee et al. [23]. They found that polysaccharides and protein that have larger size of MW and lower UV to HPSEC-DOC/UV response, to significantly fouled their low pressure (MF/UF) membrane. Polysaccharides are aldehydes derivatives of high polyhydric alcohols, which have neutral characters that can cause adsorption on a charged hydrophobic UF membrane [59]. They suggested that the neutral NOM fractions with larged sized were the prime foulants rather than humic substances. Speth et al. [24] in their study also found hydrophilic neutrals fouled more than hydrophobic acids. It can be reasoned that since polysaccharides (hydrophilic neutral) has a bulky macromolecular shape and with no electrostatic effect would definitely be prone to foul and adsorb on membrane surface. Lin et al. [25] performance a study on the effect of fractionated NOM onto a negatively charge UF membrane and observed that both larged-sized molecules of hydrophobic and hydrophilic of NOM components caused worsen flux decline. However the hydrophilic fraction was found to induce the worst fouling. Jarusutthirak et al. [26] in their study on the effect of effluent organic matter for UF membrane also found that the high molecular weight of hydrophilic component was responsible as the prime contributor of NOM fouling. It can be claimed NOM fouling was a result of low UV absorbing compounds and high molecular weight hydrophilic components that occurred through adsorption mechanisms. Inorganic particles can also affect the fouling behaviours of the organic substances. Presence of inorganic particles such as clay minerals in the surface water incurred a significant influence of competition between NOM and inorganic particles to adsorb onto the membrane surface or in the pores. High surface areas of inorganic particle enhance adsorption of organic substances on clays minerals and affect the fouling characteristic. This results in either enhancing particle deposition on the membrane or would decrease sorption of NOM onto the membrane and hence increase the membrane permeability. But opposite finding was experienced by Hong et al. [27] when studied the effect silica (inorganic particle) on fouling of zirconia tubular membrane.

3. Common Techniques for foulants identification

Researcher had used a number of techniques including inline Attenuated total reflection (ATR) Fourier transform infrared (FTIR) spectrometry, UV₂₅₄, SEC-DOC, DOC fractionation, pyrolysis-GC/MS, UF fractionation, SEM, SEM with Dispersive Spectrometer (EDS) and AFM to analyze and characterize membrane foulant. ATR-FTIR spectrometry can provide insight of foulant nature in the membrane texture that appears to be a valuable tool for foulants autopsy [28]. FTIR can also be used to determine the functional groups of certain unknown foulants which corresponding to their vibrational energy of atomic bonds. Different functional group would absorb energy at different specific wavelength that latter can be translated in intensity

response. Frequent absorption bands seen are shown in Table 2.

Table 2. Common IR spectra for humic substances, polysaccharides and proteins

Bands (cm ⁻¹)	Functional group
Humic Substances	
2940-2900	Aliphatic C-H stretching
1725-1720	Carboxylic acids
1660-1630	C=O stretching of amide group
1620-1600	Aromatic C=C
1590-1517	COO ⁻ , N-H deformation
1460-1450	Aliphatic C-H
1400-1390	OH deformation, C-O stretching of phenolic OH
1280-1200	C-O stretching, OH deformation of COOH
1170-950	C-O stretching of polysaccharide
Polysaccharides Group	
3400	Alcohol (1,2,3, Ar)
2940	Alkane
1480	Alkane
1370	1370 (starch)
1170	Tertiary alcohol
1120	Secondary alcohol
1040	Aliphatic ether
1000	Primary alcohol
775	Ethyl

But this method may also be insignificant in identifying of certain functional groups when the absorption reading gives broad overlapping bands. This phenomenon occurred due to heterogeneity of natural waters. Researchers [29] use UV₂₅₄ absorbance at 254 nm to measure the permeate and retentate of rejected compounds specifically for humic substances. The presence of unsaturated compounds would generally produce a distinct colour and can therefore be detected by UV-Vis [30]. UV₂₅₄ absorbance is sensitive to aromatic components and is an indicator for both humic acid and fulvic acid presence. Sample will first be filtered through 0.2 µm to remove particulate matter and DI is used as a blank. Difference in reading of UV₂₅₄ absorbance between feed and permeate indicates the quantity of rejected humic substance by the membrane. SUVA or specific ultraviolet absorbance is a ratio of UV at the wavelength of 254 nm and DOC. High SUVA means high aromaticity or hydrophobicity of sample in that limited DOC. Permeate from membrane filtration process which is found with high SUVA value is conforming that most of the rejected compounds are non-humic and that resulted in high value of UV₂₅₄ in the permeate. Previous studies

reported that SUVA of NOM from natural waters or ground waters was in the range of 2.4 – 4.3 L/mg-m to 4.4 – 5.7 L/mg-m respectively [31]. The molecular weight distribution of NOM was normally determined using high performance liquid chromatography (HPSEC) with online UV and DOC detection [32]. The HPSEC contains a porous gel that allows separation of molecules based on their mass and MW. Smaller molecules will access most of the pore volume while larger molecules that cannot pass the pores will be eluted first. Subsequently non-humic compound with large molecular weight such as carboxylic protein and polysaccharides will exhibit significant DOC peaks but with low area of UV₂₅₄ peak. On the other hand humic fraction such as humic acid and fulvic acid will exhibit high peaks with molecular mass between 500 daltons to 2000 daltons with high UV response. The most common technique for isolation of NOM fractions are gel filtration, ultrafiltration and adsorption using non-ionic macro-porous ion exchange resins DAX-8 and XAD-4 [33]. The surface water is fractionated into hydrophobic (HPO) which is DAX-8 adsorbable, transphilic (TPI) fraction which is XAD-4 adsorbable and hydrophilic (HPI) components which pass through the DAX-8 and XAD-4 resin without any adsorption. Application of pyrolysis and GC/MS tool as an analytical method to characterized complex organic matter had been successfully employed by Speth et al. [34] and Jarusutthirak et al. [35]. A recent study done by Speth et al. [36] using pyrolysis with GC/MS showed that hydrophilic fraction of NOM were the major foulants for a river water filtration. Pyrolysis-GC/MS method was developed by Bruchet et al. [37]. This tool is useful for characterizing NOM in terms of biopolymers such as polysaccharides, polyhydroxyaromatics, amino sugar and protein. When pyrolysing refractory compounds would release volatile fragments, which are separated and analyzed by GC/MS. Those fragments are then characterized by relative percentages according to their biopolymer. However pyrolysis -GC/MS technique is considered as a semi quantitative technique due to variation of fragments characteristics with their biopolymer structures. Beside that almost 50% of the total peaks would be classified as miscellaneous and the standards used also may not be effectively representing the classes of required compounds.

4. Fouling Mechanisms in Surface Water

The extent of rejection of solutes by membrane is the most critical parameter in membrane filtration. For a clean membrane the extent of rejection is largely influenced by pore size whereas for a fouled membrane it is determined by the electrostatic interactions between the solute and membrane [39]. Fouling of membrane is likely to happen in many instances due to a number of mechanisms such as pore blocking of solutes, cake deposition and precipitation of inorganic and organic particles at the membrane surface. Bowen et al. [40] elucidated the consecutive steps of membrane blocking in flux decline during MF as follows: (i) the smallest pores are blocked by all particles arriving to the membrane; (ii) the inner surfaces of bigger pores are

covered, (iii) some particles arriving to the membrane cover other already arrived particles while others directly block some of the pores, and (iv) a cake starts to be built. NOM fouling mechanisms on membrane processes are different and are dependent upon membrane types. For MF, pore plugging, pore blockage and cake formation were found responsible for fouling that reduces pore size and increases rejection. In case of UF, internal pore adsorption reduces the internal pore diameter and enhances rejection while in NF the fouling mechanism is mostly governed by cake deposition and concentration polarization. Many researchers suggest humic substances play a vital role in irreversible fouling of membranes. Maartens et al. [41] claimed UF membranes can remove NOM from natural brown water up to 98% but this progress impacted in flux decline in permeate volume that was due to irreversible fouling mechanism. Hydrophobic interaction between the hydrophobic NOM fraction and a hydrophobic membrane may cause more flux decline than that of hydrophilic membrane. NOM with variety of organic fractions of different hydrophobicity, hydrophilicity, molecular weight, sizes and charge densities would give different interactions in membrane filtration. Yuan and Zydney [42] have found that NOM adsorbed both inside the pores and on the membrane surface to form a cake layer. A cake layer formation is generally known to occur during surface water filtration using tight UF, NF and RO while pore blockage or direct adsorption is usually happened when using the MF. A cake layer formation is caused by electrochemical interaction and the degree of accumulation is depending on a balance between convective transport of solutes towards the membrane and back diffusion transport. Transport of large particle by drag force (convective force) is governed by an orthokinetic mechanism (inertial lift and shear induced diffusion). Inertial lift induced by wall effect tends to reduce larger particle to the membrane especially at high CFV [43]. Furthermore the shear-induced diffusion is found to increase back transport of particles. Both of inertial lift and shear-induced diffusion involving backtransport are functions of particle size. The larger the particle the higher possibility it will be back transported [44]. On the other hand back transport of small particle is controlled by Brownian diffusion, which has less effect compared to inertial lift and shear induced diffusion. Subsequently large particle in cake tend to produce less resistance for the same mass of deposited. Lahoussine-Turcaud et al. [45] describes that NOM fouling was primarily governed by pore adsorption and gel formation. In their study, they experienced 25% flux reduction for the first 5 minute of 1 nm UF hollow fibre of 10 mg/L humic acid filtration and a further declination of flux (55%) after 300 minutes. They concluded the first 5 minute rapid declination was due to irreversible adsorption of humic acid foulant. The continued flux decline was claimed to cause by humic acid gel deposition (reversible fouling) by convective transport.

5.0 Conclusion

Fouling of NOM happens by many factors and mechanisms. Factor affecting NOM and membrane interactions include

NOM characteristics, operating conditions, membrane characteristics and solution chemistry. NOM fouling occurs when dissolved organic or inorganic solute adsorbs or deposits on the membrane. Adsorption mechanism happens more instantaneous and rapid compared to cake formation but depending on the membrane properties, ionic strength, pH and presence of divalent cation. Solute deposition or gel formation occurs parallel with the magnitude of a convective flux and the extent of concentration polarization. Hydrophobicity and electrostatic interactions between solute and membrane are also reported to be the dominant factors that affect the extent of NOM fouling. Presence of electrolyte composition, low pH and high ionic strength had been found to strongly enhance the degree and rate of fouling. MF and UF are drinking water treatment processes, which are particularly suitable for the removal of suspended solids and colloidal materials such as bacteria, algae, protozoa and inorganic particulates. However this filtrations types are less successful for the removal of dissolved contaminants especially NOM in the surface water. Coagulation had been introduced to address this weakness as it is proven effective for decreasing hydraulic resistance, increasing critical flux and improving NOM removal. There is still controversy over on how the NOM affects the membrane fouling mechanisms. Some studies suggested that charge interaction and adsorptive behaviour are the responsible factors that control the NOM fouling whereas others claimed convective and diffusive particle transport that mainly dominate fouling in NOM filtration. As a matter of fact many previous studies were done using various types of membranes and were operated at high fluxes in their experiment with regards to NOM fouling. These conditions would contribute to physical accumulation due to convection, diffusion and adsorptive fouling. As a result it was difficult to distinguish the dominant factor that responsible to the fouling. Hence further study needs to be carried out in order to clarify this ambiguity and help in proper selection of membrane property, membrane configuration, pretreatment and operating conditions

References

- [1] Clever, M., Jordt, F., Knauf, R., Rabiger, N., Rudebusch, R., and Scheibel, H. 2000. Process water production from river by ultrafiltration and reverse osmosis. *Desalination*, 131: 325-336.
- [2] Maartens, A., Swart, P., and Jacobs, E.P. 1998. Humic membrane foulants in natural brown water: characterization and removal. *Desalination*, 15(3): 215-227.
- [21],[36],[64] Howe, K.J., Ishida, K.P., and Clark, M.M. 2002. Use of ATR/FTIR spectrometry to study fouling of microfiltration membranes by natural waters. *Desalination*, 147: 251-255.

- [4],[41] Maartens, A., Swart, P., and Jacobs, E.P. 1999. Feed-water pretreatment methods to reduce membrane fouling by natural organic matter". *J. Mem. Sci.*, 162: 51-62.
- [5],[22] Fan, L., Harris, J.L., Roddick, F.A., and Booker, N.A. 2001. Influence of the characteristics of natural organic matter on the fouling of microfiltration membranes. *Wat.Res.*, 35 (18): 4455-4463.
- [6] Combe, C., Molis, E., Lucas, P., Riley, R. and Clark, M. 1999. The effect of CA membrane properties on adsorptive fouling by humic acid". *J. Mem. Sci.*, 154 : 73-87.
- [7],[43] Wiesner, M.R. and Aptel, P. Mass transport and permeate flux and fouling in pressure driven process. 1996. AWWA. *Water Treatment: Membrane Processes*. McGraw-Hill. New York..
- [8] Mallevalle, J., Anselme, C., Marsigny, O. 1989. Effect of Humic substances on membrane process, in: Suffet, I. H., MacCarthy, P. (Eds.)". *Aquatic Humic Substances*, ACS. Washington. Dc., 749-767.
- [9],[43] Kaiya, Y., Itoh, Y., Fujita, K. and S. Takizawa. 1996. Study on fouling materials in the membrane treatment process for potable water". *Desalination*, 106: 71-77.
- [10],[44] Mo, L., and Huang, X. 2003. Fouling characteristics and cleaning strategies in a coagulation-microfiltration combination process for water purification. *Desalination*, 159:1-9.
- [11],[14] Nilson, J.A. and DiGiano, F.A. 1996. Influence of NOM composition on nanofiltration". *J. AWWA.*, 88(5): 53-66.
- [12] Thurman, E. 1985. *Organic Geochemistry of natural waters*", Martinus Nijhoff/Dr W. Junjk Publishers, Dordrecht.
- [13] Stevenson, F.J. 1982. *Humus Chemistry*. John Wiley & Sons. New York.
- [15] Jones, K.L. and O' Melia, C., 2000. Protein and humic acid adsorption onto hydrophilic membrane surfaces: effect of pH and ionic strength". *J. Mem. Sci.*, 165: 31-46.
- [16],[42] Yuan, W. and Zydney, A..L. 1999. Humic acid fouling during Microfiltration". *J. Mem. Sci.*, 157: 1-12.
- [17] Schafer, A.I., Schwicker, U., Fisher, M.M., Fane, A.G., and Waite, T.D. 2001. Microfiltration of Colloids and Natural Organic Matter". *J. Mem. Sci.*, 171:151-172.
- [18],[45] Lahoussine-Turcaud, V., Wiesner, M.R., and Bottero, J.Y. 1990. Fouling in tangential-flow ultrafiltration: the effect of colloids size and coagulation pretreatment". *J. Mem. Sci.*, 52:173.
- [19],[25] Lin, C., Lin, T., and Hao, O.J. 1999. Effects of humic substance characteristics on UF performance. *Water Res.*, 34(4): 1097-1106.
- [20],[21] Carroll, T., King, S., Gray, S.R., Bolto, B.A. and Booker, N.A. 1999. The fouling of microfiltration by NOM after coagulation treatment. *Wat. Sci.*, 34 2861-2868.
- [23] Lee, J., Lee, S., Jo, M., Park, P., Lee, C., and Kwak, J. 2002. Effect of coagulation conditions on membrane filtration characteristics in coagulation-microfiltration". *Environ. Sci. Technol.*, 34:3780-3788.
- [24],[34],[36] Speth, T., Guses, A.M. Summers, R.S. 2000. Evaluation of nanofiltration pretreatments for flux loss control. *American Water Works Assoc. Mem. Tech. Conf. Long Beach. CA.*, 7(2).
- [26],[35] Jarusutthirak, C., Amy, G., and Croue, J.P. 2002. Fouling characteristics of wastewater effluent organic matter (EFOM) isolates on NF and UF membranes. *Desalination*, 145: 247-255.
- [27] Hong, S., and Elimelech, M. 1997. Chemical and physical aspects of natural organic matter (NOM) fouling of nanofiltration membrane". *J. Membrane Sci.*, 132: 152-181.
- [28] Howe, K.J., Ishida, K.P., and Clark, M.M. 2002. Use of ATR/FTIR spectrometry to study fouling of microfiltration membranes by natural waters". *Desalination*, 147 251-255.
- [29] Sakol, D., and Konieczny, K. 2003. Application of coagulation and conventional filtration in raw water pretreatment before microfiltration membrane". *Desalination*, 162: 61-73.
- [30] Maartens, A., Swart, P., and Jacobs, E.P. 1998. Humic membrane foulants In natural brown water: characterization and removal. *Desalination*, 15(3):215-227.
- [31] Krasner, S.W., Croue, J.P., Buffle, J., and Purdue, E.M. 1996. Three approaches for characterizing NOM". *J. AWWA.*, 88:66-79.
- [32] Her, N., Amy, G., and Jarusutthirak, C. 2000. Seasonal variation of nanofiltration (NF) foulants: identification and contro. *Desalination*, 132: 143-160.
- [33] Aiken, G.R., McKnight, D.M., Thorn, K.A., and Thurman, E.M. 1992. "Isolation of hydrophilic organic acids from water using nonionic macroporous resin". *Org. Geochem.*, 18(4): 567-573.

- [37] Bruchet, A., Rosseau, C., and Mallevalle, J. 1990. Pyrolysis-GC-MS for investigating high-molecular weight THM precursors and other refractory organics. *Jour. AWWA.*, 82:9: 66-74.
- [38] Gray, S.R., Ritchie, C.B., and Bolto, B.A. 2004. Effect of fractionated NOM on low pressure membrane flux declines". *Water Sci. Tech.*, Vol. 4: 189-196.
- [39] Schafer, A.I., Fane, A.G., and Waite, T.D. 2000. Fouling effect on rejection in the membrane filtration of natural waters". *Desalination*, 131:215-224.
- [40] Bowen, W.R., Calvo, J.I., and Hernandez, A. 1995. Steps of membrane blocking in flux decline during protein microfiltration. *J. Mem. Sci.*, 101: 153-165.
- [44] Chellam, S., and Wiesner, M.R. 1997. Particle back-transport and permeate flux behaviour in crossflow membrane filters". *Envi.Sci.& Tech.*, 31: 819-824.

The Characteristics of the Bus Route Length and Fleet in Jakarta

Leksmono Suryo Putranto¹

¹Department of Civil Engineering
Tarumanagara University, Jl. Let. Jen. S. Parman No. 1, Jakarta 11440, Indonesia
Tel: +62-21-5672548, Fax: +62-21-5663277, E-mail: lexy@tarumanagara.ac.id

Abstract

The public transport system in Jakarta is dominated by buses. Buses provide services for about 90% of public transport passengers. The contribution of rail transport in the public transport supply is quite limited. Therefore, whilst waiting for the development of the rail based public transport system, it is important to evaluate the performance of public bus system in Jakarta. Route length and number of buses assigned to operate on a particular route are assumed to be related one to another. In this study relationship between route length and number of buses was analyzed. A Pearson correlation coefficients between those variables were calculated for individual (small, medium, regular large, express large, express-air conditioned large) and "all" bus type. In general, there was no statistically significant correlation between route length and number of buses, except for "all" bus type from pooled analysis. Another finding was that rather than playing its role as a "door-to-door" feeder, some of the small buses have got too long route and reaching the length of the medium bus route. On the contrary, some of the medium buses routes were too short (as short as 6 km). There was substantial number of large buses serving commuter trips from and to suburban surrounding Jakarta and consequently the longest route was more than 100 km.

Keywords:

characteristics, route length, number of buses, correlation, Jakarta

1. Introduction

In a very traditional society in which types and numbers of activities are limited. The geographical scope of daily activities are within walking distances. When this society become more complex and a transport mode with higher speed is required, i.e. private car. Private car might not be affordable for everybody, therefore service available to all people is required, e.g. taxis. To accommodate more cars, wider paths are required, i.e. arterials. The use of too many passenger cars is not an efficient way to travel. The length of arterials can not be added continuously. Therefore more efficient mode of travel such as a bus which can carry more passengers is required. Further capacity improvement could be made by providing exclusive bus lane. Guided system such as light rail might provide further capacity improvement. To avoid conflicts in an intersection grade-separated paths should be provided. Finally for a metropolitan, the need for fully controlled common carrier such as subway or elevated rail is justified [1]

The public transport system in Jakarta is dominated by buses. Buses provide services for about 90% of public transport passengers. The contribution of rail transport in the public transport supply is quite limited. Therefore, whilst waiting for the development of the rail based public transport system, it is important to evaluate the performance of public bus system in Jakarta. Route length and number of buses assigned to operate on a particular route are assumed to be related one to another. In this study relationship between route length and number of buses was analyzed.

According to [2] the number of buses and the number of bus routes operated in Jakarta in 2002 were as

summarized in Table 1. It can be seen that the average number of buses operated per route was varied widely between different bus sizes. The smaller the size of the bus the higher the average number of buses per route. This suggest that the smaller the size of the bus the shorter the headway. This is quite logical, i.e. to compensate the difference in carrying capacity. However, it should be noted that the average number of buses for small buses routes was quite high. This was because the operation permit of this mode of transport is given to individual operators (not to public transport companies). Therefore, there was no integrated strategy to manage the number of buses in operation during different level of demand. As a result a lot of low occupancy small buses were disturbing through traffic by practicing illegal boarding-alighting everywhere.

Table 1. The Number of Buses and the Number of Bus Routes operated in Jakarta in 2002*

Bus Size	Bus Type	Number of Buses	Number of Bus Routes	Average Number of Buses Per Route
Large	AC	4,525	93	21
	Express		76	
	Regular		52	
Medium	-	4,978	93	54
Small	-	12,195	44	277

*Source: [2]

According to [2] the distribution of route length of buses operated in Jakarta in 2002 based on bus size was as summarized in Table 1. It can be seen that there was a significant overlapping between range of route length served by small buses and medium buses. Together with the fact that there was an inefficient operation of small buses as described in the previous paragraph, this justifies the need to convert the small buses services into medium buses services.

Table 2. The Distribution of Route Length of Buses Operated in Jakarta in 2002 Based on Bus Size

Route Length Group (km)	Route Length Median (km)	Number of Route		
		Small Bus	Medium Bus	Large Bus
4-<6	5.0	1	0	0
6-<8	7.0	1	2	0
8-<10	9.0	0	2	0
10-<12	11.0	1	2	0
12-<14	13.0	2	4	0
14-<16	15.0	6	3	0
16-<18	17.0	1	7	0
18-<20	19.0	7	4	0
20-<25	22.5	13	18	10
25-<30	27.5	4	12	14
30-<35	32.5	4	15	13
35-<40	37.5	2	10	23
40-<45	42.5	1	11	20
45-<50	47.5	0	2	27
50-<55	52.5	0	1	28
55-<60	57.5	0	0	26
60-<65	62.5	0	0	17
65-<70	67.5	0	0	12
70-<75	72.5	0	0	9
75-<80	77.5	0	0	9
80-<85	82.5	0	0	4
85-<95	90.0	0	0	4
95-<100	97.5	0	0	2
>=100	100.0	0	0	3
Total Number of Bus		43	93	221
Weighted Average of Route Length (km)		21.8	27.5	51.6

2. Approach and Methods

The data of the number of buses and the length of each bus route used for this paper was taken from [3]. It was 2001 data. There was no significant additional fleet of buses in the last 5 years. The introduction of busway recently was not adding the Jakarta bus fleet significantly since some routes which were totally or partially parallel to the busway route were asked to join the busway operation.

The buses were classified into five different types, i.e.:

- small bus
- medium bus
- regular large bus
- express-non-air conditioned large bus
- express-air conditioned large bus

A correlation analysis was carried out between the route length (X) and the number of buses (Y). There was individual analysis for each bus type and there was also pooled analysis for all bus type. In this analysis, a product moment correlation coefficient (or simply known as correlation coefficient) was calculated using equation (1)

$$r = \frac{1}{n} \sum_{i=1}^n \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{S_x S_y} \quad (1)$$

where

\bar{X} is the mean value of X

\bar{Y} is the mean value of Y

S_x is the standard deviation of X

S_y is the standard deviation of Y

It was also important to evaluate the significance level of the calculated correlation coefficients. In this study a significance level of 0.05 was used.

3. Results

Number of small bus routes (in 2001) analyzed was 49. This is larger than the number of routes of small buses in 2002 (Table 1). Although there might be some route closures in 2002, but this is a bit peculiar since the 49 routes were only cover Mikrolet company which was just one of the small buses services providers. Only 43 of 93 medium buses routes were available to be analyzed, i.e. from Kopaja, Koantas Bima and Kopaja companies. Only 39 of 52 regular large bus routes were available for analysis, i.e. from PPD and Mayasari Bakti companies. There were 65 of 76 express-non air conditioned large bus routes available for analysis, i.e. from PPD and Mayasari Bakti companies. Finally, there were 50 of 93 express-air conditioned large bus routes available for analysis, i.e. from PPD and Mayasari Bakti companies.

Table 3 shows results of the correlation analysis between route length and number of buses operated in Jakarta in 2001.

Table 3. Correlation Analysis between Route Length and Number of Buses Operated in Jakarta in 2001.

Bus Type	r	Significant at $\alpha=0.05$?
All	-0.34	yes
small	-0.15	no
medium	0.04	no
regular large	0.04	no
express-non-air conditioned large	0.12	no
express-air conditioned large bus	-0.04	no

4. Discussion

It can be seen from Table 3 that in most cases there were no significant correlation between route length and number of buses operated in Jakarta in 2001. The only significant correlation was for pooled analysis of all bus. The negative sign of the correlation coefficient suggests that the longer the bus route, the lower the number of bus assigned. This was quite logical since from Tables 1 and 2 we found that larger buses with longer average route length were consisted of smaller fleet. However this was not true for individual bus type. Therefore within each bus type there was no significant relationship between route length and number of bus assigned.

5. Conclusion

For pooled analysis of all bus, the longer the bus route, the lower the number of bus assigned. This was not true for individual bus type. Another finding was that rather than playing its role as a as a “door-to-door” feeder, some of the small buses have got too long route and reaching the length of the medium bus route.

On the contrary, some of the medium buses routes were too short (as short as 6 km). There was substantial number of large buses serving commuter trips from and to suburban surrounding Jakarta and consequently the longest route was more than 100 km.

Acknowledgments

I would like to thank Eric Antonius Lung, Anthony Chandra, Teddy Maryadi, Dani Jamaludin, Reza Sangguardi, Syly, Robby Sucipto, Nicco, Anastasia Dyah, Yudhi Susanto and Dedy Setio Intan the students Civil Engineering Department, Tarumanagara University, Jakarta who helped me to collect the data for this paper.

5. References

- [1] Vuchic, V.R. 1981. *Urban Public Transportation*, Englewood Cliffs: Prentice-Hall, Inc.
- [2] Decree of the Governor of Jakarta No. 84. 2004. *The Macro Transport Pattern of Jakarta*. Jakarta: Jakarta Transport Agency.
- [3] ---. 2001. *The Jakarta Public Bus Routes*. Jakarta: Jakarta Transport Agency.

A Quantitative Approach: Information Technology Infrastructure Capability Framework in Malaysian Construction Consulting Companies

M. Z. Abd. Majid¹, Amran Rasli² and Ade Asmi³

¹Faculty of Civil Engineering, Universiti Teknologi Malaysia
E-mail: drmzaimi@yahoo.co.uk

²Faculty of Management and Human Resource, Universiti Teknologi Malaysia
E-mail: amran@ruc.dk

³Department of Civil Engineering, Universitas Indonesia
E-mail: a_asmi99@yahoo.com

Abstract

There is much information needed to manage the activities that occur throughout the construction processes. As a result, construction industries have to invest heavily in their business processes especially for Information Technology Infrastructure (ITI) capability. Being aware of the strategic impacts of information technology, many construction consulting companies have focused on improving their project performance by implementing ITI capability standard. New forms of innovative project performance, supported by ITI capability, appeared as a response to the ever-growing pressure from clients while providing new solutions to compete effectively. This paper presents the classification of ITI capability framework in Malaysian construction consulting companies based on the practitioners' perception using quantitative analysis. Finally, the classification of ITI capability framework is analysed and further grouped into two categories i.e., standardised and exclusive. One sample t-test was conducted to confirm the difference between two types of ITI capability and overall ITI capability program.

Key words: Information Technology Infrastructure Capability, Construction Consulting Companies, One Sample t-test

1. Introduction

During the past decade, Information Technology Infrastructure (ITI) capability is considered as one of the critical factors for effective business process ([1]; [2]; [3]; [4]; [5]; [6]; [7]; [8]). As the importance of the role of ITI capability in construction consulting companies increase, choosing the right of ITI capability for different project management strategies is crucial. For example, during the project management process, ITI capability is extensively utilised for planning, operating and controlling purposes.

However, the construction industry has been slow to recognize the benefits of ITI capability as a major tool beyond communication [9]. Transferring knowledge and information across projects is a major challenge for construction industries. Most construction activities are project based, characterized as short-term and task-oriented, promoting a culture where continuous learning is

inhibited. Subsequently, specialist and technical knowledge may be lost from one project to another. Gann [10] argues that ITI capability can assist the transfer of knowledge and information between project teams, enabling the development of new knowledge for innovation. Therefore, this exploratory research was conducted to classify ITI capability framework in Malaysian construction consulting companies.

2. Research framework

A research framework for ITI capability was developed as a guided for this exploratory study and it is presented in Figure 1.

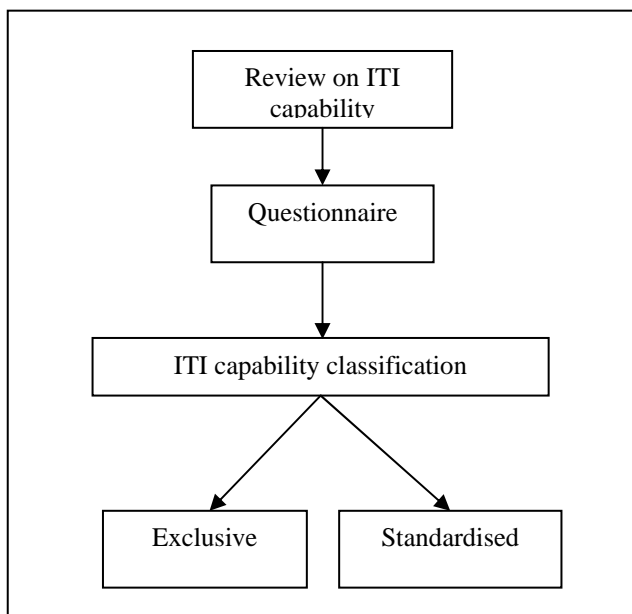


Figure 1: ITI capability research framework

Figure 1 shows that ITI capability classification has two dimensions: exclusive or standardised. ITI is “exclusive” when a company provided highly technical service and specialised in their function and operation using IT technologies. This type of ITI capability requires persons with specialised knowledge and skills to perform the job. ITI capability can be “standardised” when a company provided a relatively general technical service and non-specialised in their function and operation using IT technologies. This type of ITI capability doesn’t require persons with specific and technical skill to do job. To assess the ITI capability type, the mean total score of the ITI capability questionnaire is used. The respondents with score are less than the mean total score are classified as “standardised” and the respondent whose score are equal or greater than the total mean score are classified as “exclusive”. Furthermore, to answer the quantitative research questions and to meet the specific research objectives, analyses are made on the sample characteristics as well as the hypotheses developed on ITI capability with various statistical techniques employed. A summary of the

statistical analyses conducted are shown in the following table:

Table 1: Techniques of analysis on the data

Purpose	Technique of Analysis
Examine of the characteristics of the variables	Factor analysis Reliability analysis
Examine of the sample characteristics	Descriptive statistics
Hypotheses testing	One sample T-test

3. Demographic Findings

The nature of Information Technology can be described from a business perspective using a concept of ITI capability ([11]; [12]). Based on ITI capability research framework, a questionnaire was developed by modification from Broadbent [11] ITI capability components. Each item of the questionnaire was then anchored with a five-point Likert scale. A finalised version of the questionnaire was developed after a series of pilot testing.

One hundred and twenty two practitioners from the construction consulting companies were randomly selected to form the sampling frame. Table 2 presents the characteristics of the samples. Overall, slightly more than half (83) of the practitioners were working in civil engineering companies. The bulk of practitioners (96.7%) were from companies which are 100% owned by Malaysians. However, 76 practitioners have position as civil engineers and 15 practitioners as quantity surveyors and architects. Furthermore, 55 practitioners have working experience more than 10 years, 40 practitioners have less than 5 years working experience and 27 practitioners have 5-10 years working experience. In terms of education level, most of the practitioners have a bachelor degree. The aforementioned breakdown of respondents are reflective of the population characteristics obtained from the Construction Institute Development Board (CIDB) listing thus ensuring the sample used was a fair representation of the population parameters.

Table 2: Demographic Background of the Practitioners

Demographic Information	Frequency	Percent	Valid Percent	Cumulative Percent
Type of Company				
Quantity Surveyor	10	8.2	8.2	8.2
Civil Engineering	83	68.0	68.0	76.2
Architecture	10	8.2	8.2	84.4
Others	19	15.6	15.6	100.0
Company Ownership				
100% Malaysian ownership	118	96.7	96.7	96.7
Joint venture with foreign company	3	2.5	2.5	99.2
100% Foreign ownership	1	.8	.8	100.0
Position				
Quantity Surveyor	15	12.3	12.3	12.3
Civil Engineer	76	62.3	62.3	74.6
Architecture	15	12.3	12.3	86.9
Others	16	13.1	13.1	100.0
Total	122	100.0	100.0	
Working Experience				
Less than 5 years	40	32.8	32.8	32.8
5-10 years	27	22.1	22.1	54.9
More than 10 years	55	45.1	45.1	100.0
Education Level				
Diploma	9	7.4	7.4	7.4
Bachelor's degree	99	81.1	81.1	88.5
Others	14	11.5	11.5	100.0

4. ITI Capability Findings

Descriptive statistics based on percentages of responses for the ITI Capability related items are presented in Table 3. Upon further analysis, a preliminary conclusion could be derived that most of the respondents are

concerned about ITI Capability within their consultancy projects since most of the response for items are biased towards “quite important, Important and very important”. In most cases, a majority would rate “important” as compared to other responses.

Table 3: Descriptive Statistics of ITI Capability Attributes

ITI Capability Attributes	Very Un- important	Un-important	Quite Important	Important	Very Important
Manage communication network service (e.g., phone contact, fax, etc)	2.5%	2.5%	13.9%	27.0%	54.1%
Manage messaging services (e.g., E-mail, Notice board, etc) and business unit workstation networks (e.g., LAN/stand alone PC)	2.5%	2.5%	21.3%	35.2%	38.5%
Manage business unit workstation networks (e.g., LAN/ stand alone PC)	1.6%	7.4%	25.4%	39.3%	26.2%
Manage business applications (e.g., software applications: Microsoft Office, Microsoft Project, Primavera, etc)	1.6%	4.1%	18.9%	31.1%	44.3%
Recommend standards for IT architecture components (e.g., hardware and communications)	2.5%	5.7%	35.2%	36.9%	19.7%

Table 3: Continued

ITI Capability Attributes	Very Un-important	Un-important	Quite Important	Important	Very Important
Enforce IT architecture	2.5%	8.2%	38.5%	33.6%	17.2%
Provide multimedia operations and development (e.g., video conferencing)	2.5%	21.3%	45.1%	25.4%	5.7%
Provide intranet capability for document management	3.3%	13.1%	30.3%	35.2%	18.0%
Provide intranet capability for collaboration (e.g., local resource sharing)	3.3%	13.1%	28.7%	41.0%	13.9%
Provide electronic support for groups (e.g., documents, tutorials, CD software, etc)	3.3%	9.0%	36.1%	35.2%	16.4%
Provide data management advisory and consultancy services	0%	12.3%	36.9%	41.0%	9.8%
Manage business-unit data, including standards (e.g., oracle, database system)	0%	11.5%	50.0%	31.1%	7.4%
Manage database management system	0%	9.8%	46.7%	33.6%	9.8%
Manage, maintain and support large scale data processing facilities	0%	9.8%	44.3%	34.4%	11.5%
Provide security for firm-wide database and applications (e.g., file/data backup, anti virus software, etc)	2.5%	3.3%	22.1%	32.0%	40.2%
Implement disaster planning and recovery system for business units (e.g., backup, disk tools, etc)	2.5%	4.1%	22.1%	27.0%	44.3%
Provide technological advice and support services for internal requirement	.8%	6.6%	36.9%	41.8%	13.9%
Provide training services for new technologies (e.g., software training, etc)	1.6%	6.6%	33.6%	41.0%	17.2%
Utilise Information System (IS) for project management	1.6%	3.3%	45.1%	40.2%	9.8%
Utilise IS planning of business units	1.6%	6.6%	47.5%	34.4%	9.8%
Utilise IS to manage and negotiate with suppliers and subcontractor (by tenders, quotations in buying or dealing)	1.6%	4.9%	52.5%	32.0%	9.0%
Provide management information electronically (e.g., EIS)	1.6%	16.4%	45.1%	28.7%	8.2%
Develop and manage electronic linkages to suppliers or customers (e.g., website, E-mail address, etc)	.8%	11.5%	42.6%	32.8%	12.3%
Develop a common systems environment	4.1%	8.2%	42.6%	37.7%	7.4%

5. Factor Analysis and Reliability Test Results for ITI Capability Attributes

As shown in Table 4, the high value of 0.910 for the Kaiser-Meyer-Olkin Measure of Sampling Adequacy and the low p-value of 0.00 in Bartlett's test for sphericity indicate that the analysis is significant for subsequent factor analysis.

Table 4: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.910
Bartlett's Test of Sphericity	Approx. Chi-Square	2301.609
	df	276
	Sig.	.000

The factor analysis using Varimax with Kaiser Normalisation method was used to determine statistical validity and was able to generate five ITI capability

components as shown in Table 5. In addition, in order to ensure that the data is statistically reliable and valid, the internal consistency method was employed using reliability coefficient known as the Cronbach's alpha.

Based on Table 5, the Cronbach alpha values for the components range from 0.843 to 0.914 implied that the data is very statistically significant [13].

Table 5: ITI Capability Factors, its Components and Reliability Coefficients

No	ITI capability	Cronbach's Alpha
I	Data Management Capability <ol style="list-style-type: none"> 1. Provide data management advisory and consultancy services. 2. Manage business-unit data including standards (e.g., oracle, database system). 3. Manage database management system. 4. Manage, maintain and support large-scale data processing facilities. 5. Utilise Information System (IS) for project management. 6. Utilise IS planning of business units. 	0.914
II	Integration Capability <ol style="list-style-type: none"> 7. Manage communication network service (e.g., phone contact, fax, etc). 8. Manage messaging service (e.g., E-mail, Notice board, etc) 9. Manage business unit workstation networks (e.g., LAN/ stand alone PC). 10. Manage business applications (e.g., software applications: Microsoft Office, Microsoft Project, Primavera, etc). 11. Recommend standards for IT architecture components (e.g., hardware, operating systems, and communications). 12. Enforce IT architecture. 	0.893
III	Collaboration Capability <ol style="list-style-type: none"> 13. Provide multimedia operations and development (e.g., video conferencing). 14. Provide intranet capability for document management. 15. Provide intranet capability for collaboration (e.g., local resource sharing). 16. Provide electronic support for groups (e.g., documents, tutorials, CD software, etc). 	0.898
IV	Utility Capability <ol style="list-style-type: none"> 17. Utilise IS to manage and negotiate with suppliers and subcontractor (by tenders, quotations in buying or dealing). 18. Provide management information electronically (e.g., EIS). 19. Develop and manage electronic linkages to suppliers or customers (e.g., website, E-mail address, etc). 20. Develop a common systems environment. 	0.864
V	Security and Other Capability <ol style="list-style-type: none"> 21. Provide security for firm-wide database and applications (e.g., file/data backup, anti virus software, etc). 22. Implement disaster planning and recovery system for business units (e.g., backup, disk tools, etc). 23. Provide technological advice and support services for internal requirement. 24. Provide training services for new technologies (e.g., software training, etc). 	0.843

6. Classification of ITI capability

To assess the ITI capability types, the overall mean score of the KM questionnaire was used. The respondents whose scores are less than the overall mean

score are classified as "standardised" and the respondents whose scores are equal or greater than the overall mean score are classified as "exclusive". Table 6 shows descriptive statistic for this study.

Table 6: Descriptive Statistics

ITI capability component	N	Mean
Manage communication network service (e.g., phone contact, fax, etc).	122	4.28
Manage messaging service (e.g., E-mail, Notice board, etc) and business unit workstation networks (e.g., LAN/stand alone PC).	122	4.05
Manage business unit workstation networks (e.g., LAN/ stand alone PC).	122	3.81
Manage business applications (e.g., software applications: Microsoft Office, Microsoft Project, Primavera, etc).	122	4.12
Recommend standards for IT architecture components (e.g., hardware, operating systems, and communications).	122	3.66
Enforce IT architecture.	122	3.55
Provide multimedia operations and development (e.g., video conferencing).	122	3.11
Provide intranet capability for document management.	122	3.52
Provide intranet capability for collaboration (e.g., local resource sharing).	122	3.49
Provide electronic support for groups (e.g., documents, tutorials, CD software, etc).	122	3.52
Provide data management advisory and consultancy services.	122	3.48
Manage business-unit data, including standards (e.g., oracle, database system).	122	3.34
Manage database management system.	122	3.43
Manage, maintain and support large scale data processing facilities.	122	3.48
Provide security for firm-wide database and applications (e.g., file/data backup, anti virus software, etc).	122	4.04
Implement disaster planning and recovery system for business units (e.g., backup, disk tools, etc).	122	4.07
Provide technological advice and support services for internal requirement.	122	3.61
Provide training services for new technologies (e.g., software training, etc)	122	3.66
Utilise Information System (IS) for project management.	122	3.53
Utilise IS planning of business units.	122	3.44
Utilise IS to manage and negotiate with suppliers and subcontractor (by tenders, quotations in buying or dealing).	122	3.42
Provide management information electronically (e.g., EIS).	122	3.25
Develop and manage electronic linkages to suppliers or customers (e.g., website, E-mail address, etc).	122	3.44
Develop a common systems environment.	122	3.36
Valid N (listwise)	122	
Overall mean		3.61
Standard deviation		0.30

As shown in Table 7, the overall mean of the twenty-four ITI capability attributes is 3.61. Upon further analysis, a total of 59 practitioners were classified as “standardised” and 63 practitioners were classified as “exclusive”. The results are illustrated as shown in Table 7.

Table 7: ITI Capability Classification

Standardised (< 3.61)	Exclusive (≥ 3.61)
Total: 59 practitioners	Total: 63 practitioners
Mean: 3.14	Mean: 4.05
Standard deviation: 0.49	Standard deviation: 0.35

7. Test Between Overall ITI Capability and ITI Capability Types

A one sample T-test was used to determine whether there is difference between overall ITI capability programme and ITI capability types (i.e., standardised and exclusive). The independent sample T-test results are presented in Table 8. Based on the mean differences, it is apparent that ITI capability (standardised) has a mean below the overall mean of 3.61 as indicated by a mean difference of -0.47068. ITI capability (exclusive) has higher mean than the overall mean as shown by a mean difference of 0.44460. What is more important is that both p-values of ITI capability (standardised) and ITI capability (exclusive) are 0.000 implying there are significant

differences between ITI capability (standardised) as well as ITI capability (exclusive) and the overall mean. Thus, the null hypothesis that there is no difference between ITI

types (i.e., standardised and exclusive) and ITI capability overall programme is rejected at the 0.05 level of significance.

Table 8: Independent Sample T-Test

ITI capability type	Test Value = 3.61					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Standardised	-7.396	58	.000	-.47068	-.5981	-.3433
Exclusive	9.967	62	.000	.44460	.3554	.5338

8. Conclusion

This research was able to classify ITI capability in Malaysian construction consulting companies pursuant to practitioners’ perceptions as well as identify significant differences between the two types of ITI capability (standardised and exclusive) and the overall ITI capability program. The finalised questionnaire with twenty four items is recommended to classify ITI capability for Malaysian construction consulting companies. These items were consolidated into five ITI capability components such as: *Data Management capability; Integration capability; Collaboration capability; Utility capability and Security and Other capability.* Finally, careful interpretation is needed as to which ITI capability is better standardised or exclusive? Depending on the nature of the projects undertaken by the construction consulting companies, it would be better if parameters for adopting standardised or exclusive ITI capability be determined earlier. As such, a company can switch from standardised to exclusive ITI capability and vice versa based on the projects in hand.

References

[1] Choi, Y.S. 2000. An Empirical Study of Factors Affecting Successful Implementation of Knowledge Management. Ph.D. Dissertation, University of Nebraska.

[2] Junnakar, B. and Brown, C.V. 1997. Reassessing The Enabling Role of Information Technology in KM. *Journal of Knowledge Management*, Vol.1, pp. 29-40.

[3] Ruggles, R. 1998. The state of the notion: Knowledge management in practice. *California Management Review*, 3, Vol.40, pp. 80-89.

[4] Syed, J.R. 1998. An Adoptive Framework for Knowledge Work. *Journal of Knowledge Management*, 2, Vol.2, pp. 59 – 69.

[5] Skyrme, D.U. 1999. *Knowledge Networking: Creating The Collaborative Enterprise.* Massachusetts, Butterworth and Hainemann.

[6] Sarvary, M. 1999. Knowledge management and competition in the consulting industry. *California Management Review*, Vol.41, No. 2, pp. 95 – 107.

[7] Trussler, S. 1997. The Rules of The Game. *Journal of Business Strategy*, January/February, pp. 16-19.

[8] Zack, M.H. 1999. Managing Codified Knowledge. *Sloan Management Review*, pp. 45 – 58.

[9] Egbu, C. O., Bates, M. and Botteril, K. 2001. A Conceptual Research Framework for Studying Knowledge Management in Project-Based Environments. In *Proceeding of the International Postgraduate research Conference in the Built and Human Environment*, 15-16th March, University of Salford, UK.

[10] Gann, D. M. 2000. *Building Innovation: Complex Constructs in A Changing World.* London, Thomas Telford Publishing.

[11] Broadbent, M., Weill, P., O’Brien, T and Neo, B. S. (1996). Firm Context and Pattern of IT Infrastructure Capability. In *Proceeding of the Seventeenth International Conference on Information Systems*, Cleveland, OH, December, pp. 174-194.

[12] Weil, P and Broadbent, M. 1998. *Leveraging the new infrastructure: How market leaders capitalize on information technology.* Boston, Massachusetts: Harvard Business School Press.

[13] Nunally, J. C. (1978). *Psychometric* (2nd ed). New York: McGrawHill.

A Feature-based Music Score Recognition Method

Agus Harjoko¹ and I Gede Ngurah Agung Darmawan

¹ Electronic and Instrumentation Lab., FMIPA, Gadjah Mada University, Yogyakarta – Indonesia
Email : aharjoko@ugm.ac.id, phone +62-08164267256

Abstract

Music scores can be represented in either image format or encoded music format. Music scores stored in the encoded music format is more readily processable by a computer and takes less storage space compared to music scores stored in image format. Problem arises as how to recognize music score in image format and convert it into encoded music format.

In this research, a method for music score recognition is developed. The method starts by converting an image of music scores into the Run Length Encoding (RLE) representation of the image. Staff lines are then detected and removed from the image represented in RLE. Then image segmentation is performed using the connected component analysis method to obtain candidates of music scores. The moment invariants are computed from the music score candidates. The moment values are used to determine the similarities between the stored music score and the candidates, using city-block, Euclidean, and chessboard distance measures. If the computed similarity value is smaller than the threshold value, the corresponding music scores is said to match and the name of the music score is assigned. Otherwise the component is said to be unclassified.

Experiments with images were conducted to evaluate the performance of the recognition method. For images that is generated from a music application the method gives a success rate of 81.7% for Cityblock and Euclidean distance measures. And for images that is generated from a scanner the method gives a success rate of 46.8 % for Cityblock and Euclidean distance measures.

Keywords:

Digital image processing, music score recognition, invariant moment, connected component analysis.

1. Introduction

Computers have been used widely for object recognition. Such objects are palmprint, finger print, voice, face, characters and music scores[1,8,9,10,11,13,14]. Characters can be found in two forms: as an image of character and as an encoded character (ASCII, EBCDIC etc.). Characters stored in an encoded form can be manipulated by computers directly, such as replaced, deleted etc. Characters in an image form, on the other hand, can not be manipulated directly. It has to be recognized and encoded to be easily manipulated by a computer.

Music scores consist of head, stem and flag/hook[2]. Music scores may be found in two representations, as an image and as an encoded music scores. Encoded music scores can be manipulated by computers easily whereas music scores in an image form can't. This paper presents a method for music scores recognition based on moment invariant. A number of methods have been proposed for music score recognition. Fotinea proposed a music score recognition based on template matching and high level reasoning[3]. Fujinaga et.al. proposed an exemplar-based learning model for music recognition[4]. A method for music score recognition based on Markov source model was presented in[7]. A piano score recognition using neural

network was proposed by[10]. Once the music scores are recognized they can be stored in an encoded form thus editing of music scores can be carried out easily.

2. The Music Scores Recognition Method

The music score recognition process is discussed in this section. An object recognition process is generally achieved in four stages: preprocessing, segmentation, feature extraction and classification [5], as shown in Figure 1. In the preprocessing stage, images are prepared for segmentation. This may include noise removal, image format conversion and other image processing techniques. In the segmentation stage images are divided into objects and background. Features are extracted from objects and represented in some form in the extraction stage. These features are used for object recognition/classification in the classification stage.

The music score recognition method presented in this paper consists of four stages as shown in Figure 1. In the preprocessing stage the image of music scores are binarized and converted into the Run Length Encoding (RLE) format[5], the staff line are detected and removed. In the segmentation stage pixels are identified as objects and background and objects are labeled. Features of music scores are extracted in the extraction stage. Features of music

scores are stored and later used for recognition of music scores.

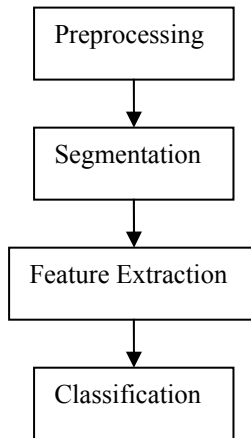


Figure 1. Stages of object recognition/ classification.

2.1 Image Acquisition

There are two types of input tested for the proposed method. First, written music scores documents. Second, captured music scores generated from a musical application.

2.2 Preprocessing

The preprocessing is intended to prepare music score images for segmentation and feature extraction of music scores. A number of steps involved in the preprocessing stage.

2.2.1 Orientation Normalization

Scanned music score documents are generally not properly aligned with the scanner. As a result the image obtained from the scanner is also not aligned and need be aligned.

The orientation normalization process was started by converting the gray level image into the binary image. It then followed by determining the orientation θ of the binary image by using [6]

$$\theta = \frac{1}{2} \tan^{-1} \frac{2\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}} \tag{1}$$

where $\mu_{p,q} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q$ (2)

being the central moment of order (p, q) , and

$$\bar{x} = \frac{1}{N} \sum_x \sum_y x, \quad \bar{y} = \frac{1}{N} \sum_x \sum_y y, \tag{3}$$

are the center of mass and N is the number of pixels in the image.

Afterwards the binary image was rotated by the amount of $-\theta$ using

$$\begin{aligned} x' &= x \cos(\theta) - y \sin(\theta) \\ y' &= x \sin(\theta) + y \cos(\theta) \end{aligned} \tag{4}$$

The result of the process of orientation normalization is the properly oriented binary image.

2.2.2 RLE Representation

In this stage the image was converted into the RLE format. The RLE format was selected for its ability to represent the image more efficiently and staff lines are readily detectable.

Run length encoding can be regarded as a data compression method where sequences of pixels of the same colors are represented as the pixel color and the number of pixels having the same color. In the case of a binary image, there are only two colors, 1 (black) and 0 (white). In this case sequences of pixels can be represented by only the number of pixels having the same color. For example, sequences of pixels (1 1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 1 1 1) can be represented as (8,9,6,5,3).

Conversion into the RLE format was achieved by scanning the image from left to right from the first row to the last row. In each scan counter for black and white pixels are stored.

2.2.3 Staff Line Detection and Removal

Staff lines are detected from the RLE representation. In the RLE representation staff lines appear as high counts. A number of parameters can be obtained from this representation, the length of staff lines, the thickness of the staff lines and the distances between two staff lines. These parameters were used for removing the staff lines. Figures 2 and 3 show the differences between an image before and after the removal of staff lines.



Figure 2. Image of music score document before staff lines were removed.

2.3 Segmentation

A connected component analysis were carried out to obtain music score candidates. In this research four and eight connectivity are used. The basic idea for connected component analysis is to obtain all connected run and assigned a label for the connected run. The run is represented in the form of a graph where each node is a run. For example, an A run for column X is a parent of a B run for the next column. Run B is a parent for a run of the next column

and so on. Children of a run were obtained by depth first traversal. For each connected component identified, a label is assigned as shown in Figure 4.



Figure 3. Image of music score document after staff lines were removed.

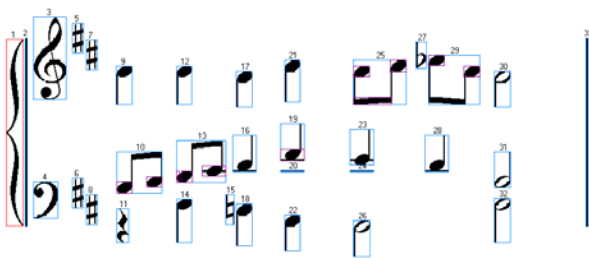


Figure 4. Labeled connected components and their bounding boxes as the results of connected component analysis.

2.4 Feature Extraction

Moment invariant [5] was selected as the features of music scores. The moment invariant was selected for its simplicity and its ability to represent rotated and un-rotated music scores as the same music score.

Moment invariant is calculated from a number of central moments and normalized

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \tag{5}$$

where $\gamma = \frac{p+q}{2} + 1, p+q = 2,3,\dots$

The seven moment invariant are

$$\varphi_1 = \eta_{20} + \eta_{02} \tag{6}$$

$$\varphi_2 = (\eta_{20} \ \eta_{02})^2 + 4\eta_{11}^2 \tag{7}$$

$$\varphi_3 = (\eta_{30} \ 3\eta_{12})^2 + (3\eta_{21} \ \eta_{03})^2 \tag{8}$$

$$\varphi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{9}$$

$$\begin{aligned} \varphi_5 = & (\eta_{30} \ 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 \\ & 3(\eta_{21} + \eta_{03})^2] + 3(\eta_{21} \ \eta_{03})(\eta_{21} + \eta_{03}) \\ & [3(\eta_{30} + \eta_{21})^2 \ (\eta_{21} + \eta_{03})^2] \end{aligned} \tag{10}$$

$$\begin{aligned} \varphi_6 = & (\eta_{20} \ \eta_{02})[(\eta_{30} + \eta_{12})^2 \ (\eta_{21} + \eta_{03})^2] \\ & + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \end{aligned} \tag{11}$$

$$\begin{aligned} \varphi_7 = & (3\eta_{21} \ \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 \\ & 3(\eta_{21} + \eta_{03})^2] + (3\eta_{12} \ \eta_{30})(\eta_{21} + \eta_{03}) \\ & [3(\eta_{30} + \eta_{12})^2 \ (\eta_{21} + \eta_{03})^2] \end{aligned} \tag{12}$$

Moment invariant is independent of translation, rotation, and scaling[5,8].

2.5 Recognition

Features of known music scores are required for recognition purpose. The reference features are extracted from known music scores as described above, i.e., through preprocessing, segmentation and feature extraction. Distances between a candidate of music score and the known music scores are calculated for classification purpose. An unknown music score is classified to a known music score if the distance to the known music scores is minimal and the distance is below a certain threshold value. If the minimum distance is larger than the threshold value the music score is said to be unclassified. This happens when a connected component consist of a number of subcomponents, see Figure 5. There are two types of components causing the connected component to be unclassified: stem_complex (of types chord and beam), and split (of types split_x, and split_y).

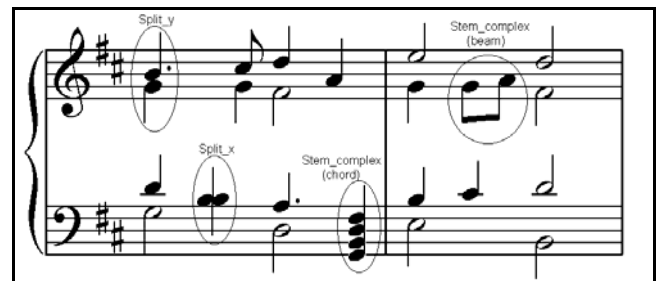


Figure 5. Image which consists of sub components (called stem_complex, split_x, and split_y).

Stem_complex

A stem_complex can be identified from its large distance from the known music scores stored in the music score database. The two types of stem-complex, chord and beam, can be distinguished from the position of the head. The processing sequence of the stem_complex is shown in Figures 6 and 7.

Split

A split component can be identified from its relatively large distance from the known music scores. A split component can be either a split_x or a split_y. In either case

the position of the head is identified. The processing sequence of split component is shown in Figures 8 and 9.

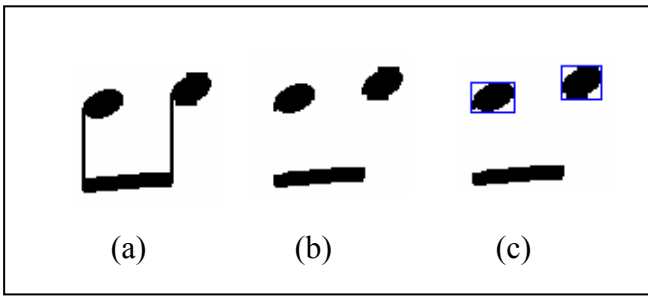


Figure 6. Processing sequence of stem_complex of type beam
a) connected component b) after removal of stem, c) after heads are found

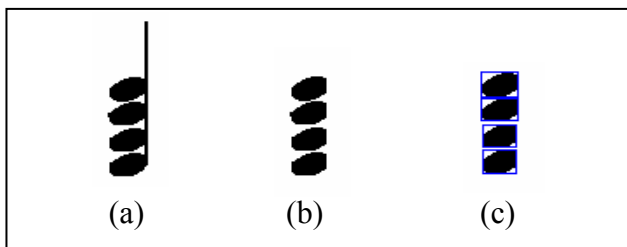


Figure 7. Processing sequence of stem_complex of type chord
a) connected component, b) after removal of stem, c) after heads are found

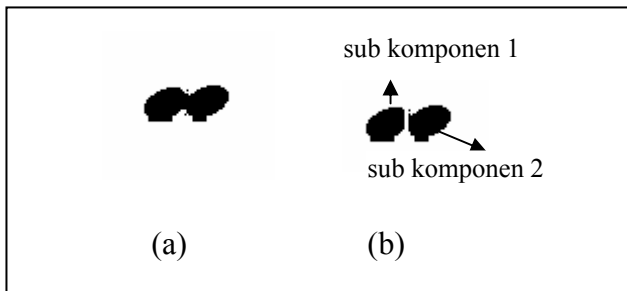


Figure 8. Processing sequence of a split_x component.

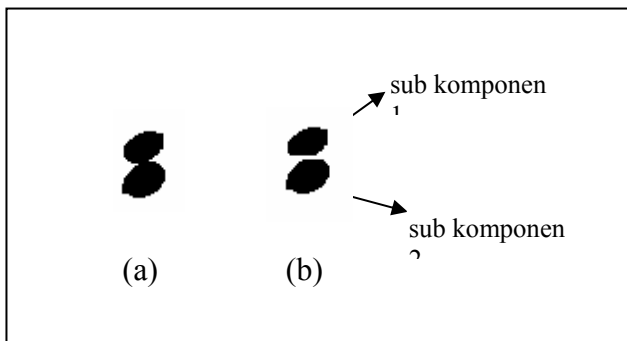


Figure 9. Processing sequence of a split_y component.

There are three distance measures used in this research. The performances of the three distance measures are compared. The three distance measures are City-block, Euclidean, and Chessboard. The three distance measures are special forms of the Minkowsky distance[12] defined as

$$d_p(x, y) = \sum_{i=1}^n |x_i - y_i|^p \quad 1/p \quad (13)$$

where x is the feature vector of the known music score, y is the feature vector of the unknown music scores and n is the length of the feature vector. The City-blok distance is a special form of (13) where ($p = 1$)

$$d_1 = \sum_{i=1}^n |x_i - y_i| \quad (14)$$

The Euclidean distance is another special form of (13) where ($p = 2$)

$$d_2 = \sum_{i=1}^n (x_i - y_i)^2 \quad 1/2 \quad (15)$$

The Chessboard distance is another special form of (13) where ($p = \infty$)

$$d_\infty = \text{Max}_{i \in N} |x_i - y_i| \quad (16)$$

Once the music scores is known, the pitch is determined by looking at the position of the head relative to the staff line resulting in music scores with pitch a, b, c etc, as shown in Figure 10.



Figure 10. Pitch of music scores.

3 Experimental Results

Two types of input are used for testing the proposed method. The first type is a screen captured music scores generated from a music application. The staff line of this input is nicely aligned with the image space. Furthermore, the the music score components are nicely spaced. The second type of input is a scanned written music score documents. This input poses more challenges than the first one. The staff lines are generally not aligned to the image space and the music score components often touching each other.

Music scores are classified/recognized as a full not (not 1), not 1/2, not 1/4, not 1/8, not 1/16, or not 1/32. The next step is to determine the pitch by looking at the head position relative to the staff lines. The results is music scores having pitch a, b, c, d, etc as shown in Figure 10. Figure 11 shows the music score recognition system developed in this research.

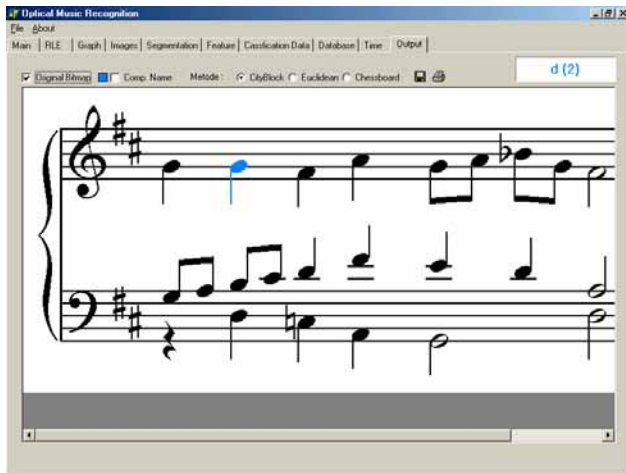


Figure 11. The music score recognition system developed in this research.

The success rate of the proposed method for the two types of input are shown in Figures 12 and 13. Table 12 show that the City-block and Euclidean distance perform equally well yielding success rate of 81.7%. The Chessboard distance gives a success rate of 63.6%. For input obtained by scanning written document, both City-block and Euclidean distances give a success rate of 46.8% whereas the chessboard distance resulted in a success rate of 42.5%. Most of the problems in recognizing music scores are in the preprocessing and segmentation stages.

Table 1. Classification results obtained from type one input (screen captured image generated from a music application).

	# scores	CityBlock	Euclidean	Chessboard
		# recognized (%)	# recognized (%)	# recognized (%)
1	33	27 (81,8)	27 (81,8)	22 (66,7)
2	38	31 (81,6)	31 (81,6)	23 (60,5)
Total	71	58 (81,7)	58 (81,7)	45 (63,6)

Table 2. Classification results obtained from type two input (scanned from written music score document)

	# scores	CityBlock	Euclidean	Chessboard
		# recognized (%)	# recognized %	# recognized %
3	33	17 (51,5)	17 (51,5)	15 (45,5)
4	38	16 (42,1)	16 (42,1)	15 (39,5)
Total	71	33 (46,8)	33 (46,8)	30 (42,5)

4. Conclusion

From this research we conclude that moment invariant can be used for music score recognition. In this research three distance measures were used, i.e., City block, Euclidean and Chessboard distance measures. The City block and Euclidean distance measures give the same success rate of 81.7% for music scores generated using music application and 46.8% for input scanned from music score document.

The Chessboard distance gives a success rate of 63.6% for music scores generated using music application and 42.5% for input scanned from music score document.

5 References

- [1] Daugman, J. 1993. How Iris Recognition Work. *IEEE Transaction On Circuits And Systems for Video Technology*, Vol 14, No 1, page: 21-30.
- [2] Elliott, R. 1999. *Fundamentals of Music*, Prentice Hall
- [3] Fotinea, S.E. 1999. An Optical Notation Recognition System for Printed Music based on template matching and high level reasoning, URL <http://www.eng.cam.ac.uk/~ajr/esca99/Hirschman.pdf>
- [4] Fujinaga, I., Moore, S., and Sullivan, D. S. 1998. Implementation of exemplar-based learning model for music recognition, URL <http://gigue.peabody.jhu.edu/~ich/research/icmpc98/icmpc98.exemplar.pdf>
- [5] Gonzales, R.C and Woods, R.E. 1992. *Digital Image Processing*, Addison-Wesly Publishing Company.
- [6] Jain, A.K., 1989, *Fundamental of Digital Image Processing*, NJ, Prentice Hall Information and System Sciences Series.
- [7] Kopec, G.E. and Chou, P.A. 1995. Markov Source Model for Printed Music Decoding, URL <http://www2.parc.com/istl/members/kopec/jei-1-96-music.pdf>.
- [8] Lu, G. 1999. *Multimedia Database Management System*. Artech House
- [9] Masek, L. 2003. Recognition of Human Iris Patterns for Biometric Identification. Thesis. School of Computer Science and Software Engineering, University of Western Australia.
- [10] Miyao, H. and Nakano, Y. 1996. Note Symbol Extraction for Printed Piano Scores Using Neural Networks. URL <http://sfbc.uni-dortmund.de/home/English/Publications/Reference/Downloads/KKR02.pdf>
- [11] Putra, I.K.D, Susanto A., Harjoko A., Widodo, T. 2004. Identifikasi Telapak Tangan dengan Memanfaatkan Alihragam Gelombang Singkat, *PAKAR*, Vol.5, No.3, UTY, Yogyakarta.
- [12] Schalkoff, R., 1992, *Pattern Recognition, Statistical, Structural, and Neural Approaches*, Jhon Wiley & Sons, Inc.
- [13] Selzer, J. 2005. An Implementation of Local Feature-based Iris Recognition. Thesis. Computer Science Department, University of Alberta, Canada.
- [14] Willits A.E., 2003, *Low Resource Speaker Identification*, M.Sc. Thesis, The Department of Theoretical and Applied Linguistics, University of Edinburgh, September 2003.

Oxygen Enrichment Properties of Polyethersulfone/Polyimide Blends-Zeolite 5A Mixed Matrix Membranes

Tutuk D. Kusworo^{1,2} A.F. Ismail^{1,*} and A. Mustafa¹

¹ Membrane Research Unit
Faculty of Chemical and Natural Resources Engineering
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia

² Department of Chemical Engineering
University of Diponegoro, Semarang, Indonesia
Jl. Prof. Sudharto Tembalang, Indonesia

Abstract

We have developed polyimide-polyethersulfone (PI-PES) blends-zeolite mixed matrix membranes to investigate the separation characteristic of oxygen and nitrogen. This study investigated the effect of composition blending, membrane preparation and zeolite loading on the gas separation performance of these mixed matrix membranes. A variety of compositions of PI-PES mixed matrix membranes produced varying permeability coefficient and ideal separation factors. This indicated that the present miscible blend gave rise to decrease in the free volume by the physical interaction between two polymer components. Beside that from differential scanning calorimetry measurements of pure and PI/PES blends membranes show one unique glass transition temperature supporting the miscible character of the PI/PES mixture. Mixed matrix membrane with 20/80 % PI/PES and 25/75 % polymer-solvent produced the highest O₂/N₂ selectivity around 5.35 %.

Keywords: mixed matrix membrane, gas separation

1. Introduction

Membrane offer attractive opportunities for gas separation industries such as natural gas processing, landfill gas recovery, air separation and hydrogen recovery [1]. Glassy polymer membranes perform well in separations of mixtures of gases such as O₂/N₂, H₂/CO, and N₂/CO₂. Materials with improved anti-plasticization properties and have characteristics of both high permeability and selectivity are desirable and essential for uses in harsh environment. Good physical and gas separation properties ensure polyimides have been studied membrane material for gas separation. However, the investigation of polymer material such as polyimide for gas separation has been challenged by the upper bound trade-off limitation between the productivity and the selectivity [2]. Meanwhile, the rigid porous materials such as carbon molecular sieves and zeolites have difficulties in forming continuous and defect-free membranes of practical meaning in spite of their superior gas separation properties. Mixed matrix membranes comprising molecular sieve entities embedded in a polymer matrix have been recognized as a promising alternative to the conventional membranes.

Currently, much more interests are given to the glassy polymers [5]. Meanwhile, the major problem of mixed matrix using glassy polymer is adhesion between the polymer phase and the external surface of the particle. It seems that the weak polymer-filler interaction makes the filler tend to form voids in the interface between the polymer and filler. Therefore, the resultant membranes generally have deteriorated selectivity. Various methods addressing have been proposed to improve the polymer-filler contact. Duval et al. [6] found a poor adhesion between the polymer phase and the external surface of the filler. To overcome this problem he proposed above glass transition temperature on the casting preparation membrane. In another study, Rojey et al. [7] observed significantly enhancement for H₂/CH₄ selectivity with comprising zeolite 4A in the Ultem matrix. Süer et al. [8] reported the significantly increase the separation performance in mixed matrix membrane using polyethersulfone with zeolite 13 X and 4A. They found that both gas permeabilities of N₂, O₂, Ar, CO₂, H₂ and selectivities are increase with increasing the zeolite loadings. High amount of zeolite induces formation of microporous cavities and channelling, demonstrating the weak interactions and incompatibility of these materials. As a tool to improve zeolite incorporation in polyimide, silane coupling agents, polymer coating were also made. It was reported by Vankelecom et al. [9] that mixed matrix

* Corresponding Author. E-mail: afauzi@utm.my

membrane with silylation zeolite in polyimide using borosilicate exhibited small improvement of the polymer-filler contact. Some studies have suggested selection criteria of some material and preparation protocols in order to match the necessary transport characteristics of materials to form high separation performance of mixed matrix membrane [10, 11]. Vu et al. [12] investigated the mixed matrix membranes using carbon molecular sieve (CMS) particles as inorganic filler. The CMS was incorporated into commercial glassy polymer Ultem and Matrimide 5218. They reported that mixed matrix membrane using CMS particles showed enhancements by up to 40-45 %, respectively in CO₂/CH₄ selectivity over the intrinsic selectivity of the pure Ultem and Matrimide polymer. They used polymer coating on molecular sieve to remove the interface defect; therefore separation performance of mixed matrix membrane could be improved. Anson et al. [13] studied the effect of activated carbon as inorganic filler in ABS copolymer mixed matrix. They found that the gas separation performance showed increase with increasing loading activated carbon up to 40 %.

The present work proposes an alternatively preparation of mixed matrix membranes for O₂/N₂ separation by using zeolite as inorganic filler. The zeolite was incorporated into blending of polyimide and polyethersulfone as polymer matrices. Polymer blending is a possibility of modifying material properties. A blend can show new properties not found for single polymers. Furthermore, blending is an option for reducing the price of membranes. In this contribution, the permeation characteristics of O₂/N₂ in polyimide and polyethersulfone blends-zeolite mixed matrix membrane are presented. The observed permeation behaviour was analyzed with the free volume theory.

2. Methods

2.1. Material

Radel Polyethersulfone from Solvay Advanced Material and Polyimide resin from Alfa Aesar were the polymers used in this study. The polymers were dried in a vacuum oven at 120 °C for overnight before dope preparation; N-methyl-pyrrolidinone (NMP) from Merck was used as the solvent due to its low toxicity. The inorganic filler molecular sieves involved were zeolite 5A purchased from Aldrich and the particle size is about 4 µm. in order to remove the adsorbed water vapour or other organic vapors all zeolites were dehydrated at 300 °C for 3 h before use.

2.2. Dope preparation

The homogeneous polyimide and polyethersulfone were prepared according to the following procedure. The inorganic molecular sieve particles were dispersed to the solvent and stirred for 24 hours followed by the addition of a desired amount of polyimide. The solution was agitated with the Kika stirrer at a low speed about 30 rpm at least 24 hours for the polymer to dissolve completely.

After the solution homogenously, a desired amount of polyethersulfone was added to the solution. All the solutions were agitated at low speed at least 2 day until the polymer to dissolve completely and homogenously.

2.3. Membrane preparation procedures

The casting process was performed by using a pneumatically controlled casting machine. The casting solution was cast on a glass plate with a casting knife gap setting of 150 µm. Based on the literatures [5, 11] high processing temperatures close with the *T_g* during membrane formation was applied to eliminate the void between polymer and filler.

2.4 Gas permeability measurement

The gas permeation properties of the flat mixed matrix membrane and pure polymer film were measured by the variable volume method, using an upstream pressure up to 8 bars, while the downstream pressure was effectively atmospheric. The high pressure gas controlled by a line regulator is fed into the test cell. Permeation experiments were carried out at room temperature. The experiments are performed after a period of equilibrium of 15 min. A schematic of the gas permeability equipment is shown in Figure 1. The material used for the construction was stainless steel 316 and double walled permeation cells. The permeate volumetric flow rate was measured by means of a soap bubble meter reading to 0.05 cm³. The effective area inside the cell was 14.5 cm². Two O-rings were used as the sealing agent. The downstream side was always purged with the test prior to the permeation test. The permeance or pressure-normalized flux is expressed as a thickness normalized permeation rate, *P/l*, and was calculated as

$$\frac{P}{l} = \frac{Q_i}{(\Delta p)(A)} \quad (1)$$

where *Q_i* is the volumetric flow rate of gas “i” at standard temperature and pressure, *Δp* is the transmembrane pressure difference and *A* is the membrane active surface area. Permeances are expressed in gas permeation units, barrers.

2.4. Characterization of mixed matrix membrane

Several characterization techniques were performed on the mixed matrix membranes. The MMM were analysed wide SEM, FTIR and DSC. Membrane morphology was determined with a Philip XL-40 scanning electron microscope (SEM). Cross-sections of the membranes were obtained by freeze-fracturing by the immersion of the sample in liquid nitrogen. The membranes were mounted on aluminium disk with double surface tape. The sample holder was then placed and evacuated in a

sputter-coater with gold. The changes in the chemical structure during the blending were followed using Fourier transform infrared spectroscopy (FTIR). The IR absorption spectra were measured at room temperature from 4000 to 500 cm^{-1} with a spectral resolution of 8 cm^{-1} and averaged over 16 scans. The glass transition temperature of each cast film was analysed using differential scanning calorimetry on Mettler Toledo DSC 822e. A small piece of membrane or pure polymer was first stored under vacuum at 100 °C for 24 hours to remove adsorbed water; then weighed and placed into aluminium DSC pans. The scanning range was 50- 320 °C with scanning rate of 10 °C min^{-1} in the first DSC cycle to remove thermal history and then cooled from 320 to 25 °C at the rate of 10 °C/min; finally carried out with the same procedure in second cycle.

3. Result and Discussion

3.1. Characterization of mixed matrix membrane

The cross section SEM micrographs of PI-PES blends-zeolite 5A mixed matrix membrane are presented in Figure 1. These micrographs demonstrate a better distribution of zeolite as well as very good polymer-filler contact. The result indicated that combination of vacuum degassing, mixing time and low speed agitation can effectively avoid the formation of voids between polymer and zeolite. The results indicated that the application of close T_g temperature for preparation of mixed matrix membrane could be eliminate the void between polymer and inorganic as a filler. The result is also agreement with previous study by Li et al [5].

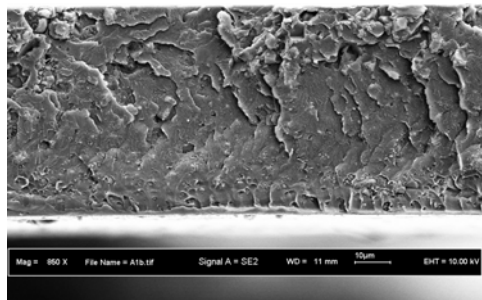


Figure 1: Cros section SEM image of mixed matrix membrane PES/Polyimide (20/80)-Zeolite 5A

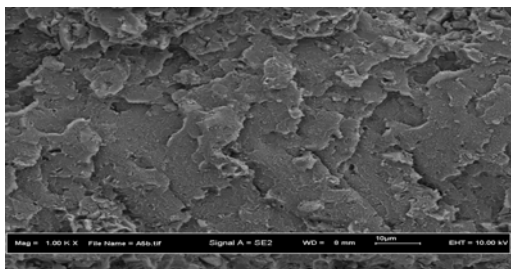


Figure 2: Cros section SEM image of mixed matrix

membrane PES/Polyimide (50/50)-Zeolite 5A

The glass transition temperature of the homopolymer and their blends is presented in Table 1. The T_g for each membrane was determined from the heat flow versus temperature curve using onset method. Measurement of T_g is one of the easiest determination of whether they are miscible or immiscible in polymer mixtures. A single T_g between individual two polymer appears for miscible blends whereas two T_g s are shown for immiscible blends. From DSC scans of PI-PES blends-zeolite mixed matrix membranes with different composition show one single glass transition temperature for each blend indicate that the PI-PES blends are miscible blends. The T_g 's of PI/PES-zeolite mixed matrix showed the linear relationship against the polymer composition. The T_g of miscible polymer blends can be correlated with the Fox equation ($1/T_g = x_1/T_{g1} + x_2/T_{g2}$) or logarithmic model ($\ln T_g = x_1 \ln T_{g1} + x_2 \ln T_{g2}$) where T_g = blend glass transition temperature, T_{g1} , T_{g2} = PI and PES glass transition temperature and x_1 , x_2 = the respective weight fraction of polymer 1 and 2 in the blend [13]. As shown in Figure 2, the experimental T_g values are generally in good agreement with those predicted by the logarithmic model while at low polyimide concentration (< 20 %), they are a good agreement with the Fox model. There is a deviation between the experimental results and model equation values when the polyimide concentration is higher than 40 wt %. It may imply non-fully miscibility of these two polymers when polyimide concentration is higher than 40 wt %. The T_g of the mixed matrix membrane increased with the incorporation of the zeolite particles. However, the increasing of zeolite loading was a gradually or slightly increased the T_g of mixed matrix membrane about 1.5-3 °C (See Figure 3). Increases in T_g maybe caused by zeolite was grown in the polymer solution and the high temperature processing will be formed cross linkage between the inorganic phase and the organic polymer matrices. The physical properties of membrane was changed with increasing zeolite loading such as mixed matrix membrane more fragile, brittle therefore decrease the tensile strength compared to low zeolite loading. This phenomenon can be explained by the creation of voids in the membrane occurring at the zeolite-membrane interface. Because of the difference of the chemical nature, the adhesion between the zeolite particles and the polymer matrix (polyimide and polyethersulfone) is probably weak. Perhaps, even some of zeolite aggregates appear at higher loadings. The experimental results showed that the presence of inorganic or molecular sieving on the polymer matrix can affect the fundamental properties (e.g. T_g , tensile strength) of the mixed matrix membrane.

Table 1: Glass transition temperature of PI/PES-zeolite 5A

mixed matrix membrane

PI/PES (%w/w)	T_g (°C)
0/100	220.75
20/80	223.14
30/70	232.43
50/50	267.34
80/20	300.13
100/0	327.12

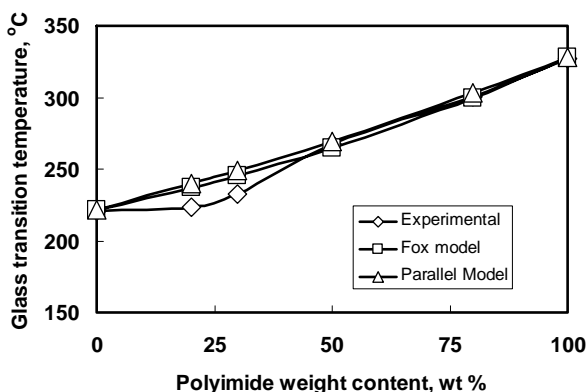
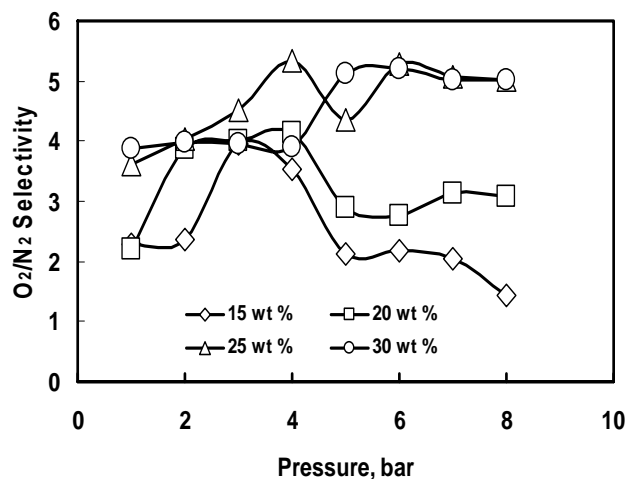


Figure 3: Glass transition behaviour of PES/PI- Zeolite 5A mixed matrix membrane

3.2 Gas permeation characteristic

Figure 4 shows the effect of polymer concentration of PES/PI blends-Zeolite 5A mixed matrix membrane on the oxygen and nitrogen selectivity characteristics. The selectivity increases with increasing polymer concentration. The higher polymer concentration in the casting solution will produce the mixed matrix membrane with small free volume. Therefore, selectivity will increase as polymer concentration increase. The results were in agreement with previous study by Kim et al (1997).

Figure 4: Selectivity of O₂/N₂ of PES/PI- Zeolite 5A mixed matrix membrane for various polymer concentrations

4. Conclusion

The permeation and selectivity characteristics were investigated for PES/PI blends-zeolite 5A mixed matrix membrane. The selectivity was significantly affected by polymer concentration and blends composition. The miscible of PES/PI was confirmed by the single T_g for each composition. The PES/PI polymer blends-zeolite 5A mixed matrix can be viewed as a new economical high performance raw material suitable for gas separation membranes

Reference

1. Kim, Y.K., H.B. Park and Y.M., Lee, "Carbon molecular sieve membranes derived from thermally labile polymer containing blend polymers and their gas separation properties", *Journal of Membrane Science*, 243. (2004) 9-17.
2. Jiang, L.Y., T.S. Chung., and S. Kulprathipanja, "An investigation to revitalize the separation performance of hollow fibers with a thin mixed matrix composite skin for gas separation", *Journal of Membrane Science*,
3. M. Jia, K.V. peinemann, R.D., Behling, Molecular sieving effect of zeolite filled silicone rubber membranes, *Journal of Membrane Science*, 57. (1991) 289-296.
4. J.M. Duval, B. Folkers, M.H.V. Mulder, G. Desgrandchamps, C.A. Smolders, Adsorbent-filled membranes for gas separation. Part 1. Improvement of the gas separation properties of polymeric membranes by incorporation of microporous adsorbents, *J. Membr. Sci.* 80 (1993) 189-198.

5. Y., Li, T.S. Chung., C., Cao, and S. Kulprathipanja, , “ The effect of Polymer Chain rigidification, zeolite pore size and pore blockage on Polyethersulfone (PES)-zeolite A mixed matrix membrane, *Journal of Membrane Science*, 260. (2005) 45-55.
6. J. M., Duval, A.J.B., Kemperman, B., Folkers, M.H.V. Mulder, G. Desgrandchamps, C.A. Smolders, Preparation of zeolite filled glassy polymer membranes, *J. Appl. Polym. Sci.* 54 (1994) 409–418.
7. A. Rojey, A. Deschamps, A. Grehier, E. Robert, Separating gases using zeolite-containing composite membranes, US Patent 4,925,459 (1990).
8. M.G. Suer, M. Bac, L. Yilmza, Gas permeation characteristics of polymer–zeolite mixed matrix membranes, *J. Membrane Sci.* 91 (1994) 77-86.
9. Vankelecom, I.F.J. S. Van den Broeck, E. Merckx, H. Geerts, P. Grobet, J.B. Uytterhoeven, , “Silylation to improve incorporation of zeolites in Polyimides Films”, *Journal of Physic Chemistry* 100, (1996) 3753-3758.
10. C.M. Zimmerman, A. Singh, W.J. Koros, Tailoring mixed matrix composite membranes for gas separations, *J. Membr. Sci.* 137 (1997) 145–154.
11. R. Mahajan, W.J. Koros, Factors controlling successful formation of mixed-matrix gas separation materials, *Ind. Eng. Chem. Res.* 39 (2000) 2692–2696.
12. D.Q. Vu, W.J. Koros, S.J. Miller, Mixed matrix membranes using carbon molecular sieves I. Preparation and experimental results, *J. Membr. Sci.* 211 (2003) 311–334.
13. S.H. Kim, D. Kim, D.S. Lee., Gas Permeation behavior of PS/PPO blends, *Journal Membrane Science* 127 (1997) 9-15

The Effect of Cross Flow Velocity on The Permeate Flux During Cyclodextrins Bioseparation

A.M. Mimi Sakinah^{1,2,3}, A.F. Ismail^{2,*}, Rosli Md Illias³, Osman Hassan⁴, A.W Zularisam²

¹*Faculty of Chemical Engineering and Natural Resources,
University of College Engineering and Technology of Malaysia (KUKTEM),
Kuantan Pahang, Malaysia
Tel: +60-095492322, Fax: +60-095492322, E-mail: mimi@kuktem.edu.my*

²*Membrane Research Unit, Faculty of Chemical Engineering and Natural Resources,
University Technology of Malaysia, 81310 Skudai Johor, Malaysia
Tel: +60-075535592, Fax: +60-075581463, E-mail: afauzi@utm.my*

³*Bioprocess Engineering Department, Faculty of Chemical Engineering and Natural Resources,
University Technology of Malaysia, 81310 Skudai Johor, Malaysia
Tel: +60-075495564, Fax: +60-075581463, E-mail: r-rosli@utm.my*

⁴*School of Chemistry and Food Science, Faculty of Science and Technology,
Universiti Kebangsaan Malaysia, 43000 Bangi Selangor, Malaysia
Tel: +60-0389213805, Fax: +60-0389213232, E-mail: osman@psikukm.my*

Abstract

Cyclodextrins (CDs) are cyclic oligosaccharides which from starch via cyclization or intramolecular transglycosylation of the cyclodextrin glucosyltransferase (CGTase). This specialty chemical have a wide range of applications in food, pharmaceutical and chemical industries, agricultural and environmental engineering. CDs can be produce by using an Enzymatic Membrane Reactor (EMR). EMR is the integration of membrane technology and an enzymatic reactor which is widely used in fermentations, enzyme-catalyzed reactions and biotransformations. Although EMR is a clean and environmental friendly technology, the application of membrane technology in EMR is still limited. In addition, flux declination is one of the significant factors that can hinder the membrane application. This declination of flux was found to reduce the membrane productivity as well as affecting its separation performance. Thus, the main objective of this research is to determine the flux declination during CDs production at various cross flow velocity (CFV). The experiment employed hollow fiber ultrafiltration membrane to separate CDs from the reaction mixture. The experimental results showed that the flux rapidly declined at low CFV. This is due to the deposition of cake on the membrane surface and within the pores matrix occurred during CDs separation process.

Keywords: Cyclodextrins; EMR; Flux declination; CFV; Hollow fiber membrane

1. Introduction

Nowadays, application of membrane in biotechnology industries is gradually emerging. The combination of membrane separation and enzymatic reactor are that is known as an enzymatic membrane reactor (EMR) has been found capable to separate enzymes from products and/or substrates by semipermeable membrane. The convective force across the membrane can selectively enhance the separation of substrate and products from the reaction mixture. The enzymes is retained within the system by the membrane, while the product passes through the membrane as a

*Corresponding Author. E-mail: afauzi@utm.my

Tel: +60-075535592, Fax: +60-075581463

permeate. The EMR present advantages such as high enzyme loads, prolonged enzyme activity, high flow rates, reductions in costs, energy and waste products by recycling, easy reactor operation and control, straightforward scale-up to large systems and high yields of pure material [1,2].

In this study, the EMR was applied to produce and separate cyclodextrins (CDs) from starch. The cyclodextrins glucosyltransferase (CGTase) and starch degradation products were retained within the system by membrane, allowing the establishment of a continuous operation with starch and enzyme feed and CDs withdrawal [2-4]. CGTase

is an enzyme which capable of converting starch and related substrate into CDs. CDs are cyclic oligosaccharide composed of α -1,4-glycosidic-linked glucosyl residues produced from starch or starch derivatives using CGTase. CDs can solublize hydrophobic materials and entrap volatile components by forming inclusion complexes with organic compounds and thus enhance their chemical and physical properties. These properties have led to the commercial application of CDs [5-6] in food, pharmaceutical, cosmetic, agricultural and plastic industries as emulsifiers, antioxidants and stabilizing agent [7]. However, the extensive use of CDs is still restricted by high production cost of CDs [8].

Although there are many advantages of membrane, the application of membrane technology in EMR is still limited. This is due to the fouling problem, which reduces the membrane performance. Fouling in membrane separation occurred when the flux decline as the function of time due to the increment of hydraulic resistance. Parallel with this scenario, this study focuses in determining the flux declination and effect of cross flow velocity (CFV) on the hydraulic resistances during the CDs separation by using EMR.

2. Materials and Methods

2.1 Cyclodextrins production

The EMR system was developed to evaluate the performance of our locally produced hollow fiber ultrafiltration membrane. As shown in Figure 1, the EMR system is comprised of an enzymatic stirred reactor with installed temperature controller, a membrane module unit, a pump, feed and retentate pressure gauges. The ultrafiltration membrane module has a length of 30 cm and a diameter of 22 mm. The locally produced PES membrane with a MWCO of 32,000 Dalton was used in this experiment. The membrane diameter and effective area are 600 μm and 0.027 m^2 , respectively. The enzymatic reactor consists of a stainless steel vessel with a mechanical stirrer attached. This vessel was filled with 2 % of raw tapioca starch solution mixed with CGTase enzyme (200 μl /100 ml reaction volume which has an enzyme activity of 0.8 unit/ml-optimal conditions suggested by Novo Nordisk, Denmark). The reaction mixture was continuously pumped to the membrane module and recycled back to the enzymatic reactor. The temperature and pH of the enzymatic reactor was maintained at 60°C and 6.0 respectively. The operational transmembrane pressure (TMP) for membrane filtration was kept constant at 2 bars.

2.2 Resistances-In-Series Model

Resistances-In-Series model is most widely use in determination of hydraulic resistance in membrane separation. There are five parameters of membrane resistance-in-series model based on Darcy's law which were used to quantify their influences on flux decline [9]:

$$J_v = \frac{\Delta P}{\mu(r_m + r_{cp} + r_g + r_{a1} + r_{a2})} \quad (1)$$

where J_v is flux through the membrane (m/s), ΔP is the transmembrane pressure (Pa), μ is the dynamic viscosity (Pa s), r_m is the membrane hydraulic resistance, r_{cp} is the concentration polarization resistance, r_g is the gel layer resistance, r_{a1} is the weak adsorption resistance, r_{a2} is the strong adsorption resistance (all resistance are in m^{-1}). In this case, the osmotic pressure is considered into the concentration polarization. In this study, there several types of resistances contributed to both reversible and irreversible fouling mechanisms. The concentration polarization (r_{cp}) and gel layer resistance (r_g) were assumed as reversible fouling mechanisms, which could be removed by water. However, the weak and strong adsorptions were categorized as an irreversible fouling mechanism. The weak adsorption was removed by chemical cleaning, while the strong adsorption remained onto the membrane surface.

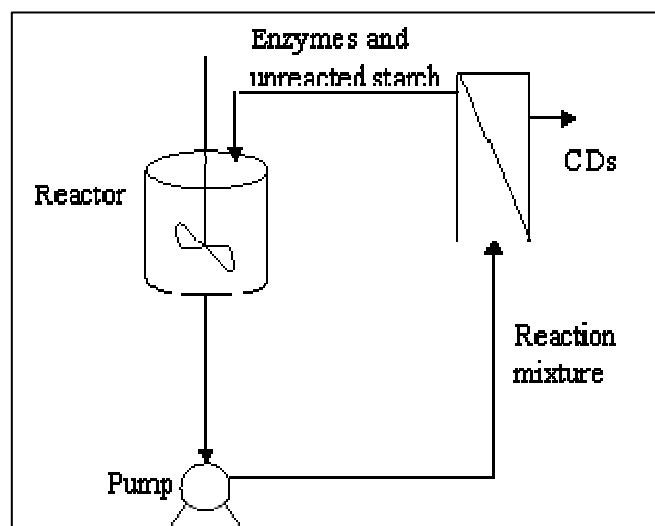


Figure 1. Schematic diagram of EMR system

The distilled water (DI) and reaction mixture were cross-flow filtered using the following procedure to obtain all hydraulic resistances quantitatively. Clean water was first filtered through the membrane to obtain the pure water flux of membrane (J_{pwp}) until a constant flux was achieved. Then, the reaction mixture was fed and the permeate rate was monitored over the time. After the permeate rate reached a constant value (that is, the permeate of fouled membrane), DI replaced the reaction mixture and the applied pressure was released to remove the concentration polarization layer. The next J_{pwp} of the membrane was taken in order to determine the concentration polarization resistance value. The fouled membrane was then rinsed with DI at higher applied pressure. This procedure was conducted to ensure the gel layer was totally removed from the membrane surface. The third J_{pwp} was taken so that r_g value could be determined. The membrane was then rinsed with 0.1M NaOH solution for 20 minutes in order to

dissolve the weak adsorption layer off the membrane surface and pores. The fourth J_{pwp} was measured and the r_{a2} were calculated using Equation 1.

3. Results and Discussion

3.1 Flux decline in separation of CDs

Figure 2 shows the declination of flux in the ultrafiltration cross flow membrane. The membrane flux was first obtained to determine the hydraulic resistance of clean membrane. Subsequently, by using the Darcy's law, the declination of flux was obtained parallel with the increasing amount of total resistances. Based on Figure 2, the permeate flux continuously decrease with time until the flux reached a steady state in 70 minutes of the separation. The permeate flux has been decreased 20% of initial permeate flux. In this stage, the cake layer has been achieved an equilibrium thickness where the hydraulic resistance of the separation was also constant. Even though the applied pressure was increased, there was still no significant change in the permeate flux at the equilibrium stage. In order to recover the initial flux, hydraulic and chemical cleaning or backwashed method must be applied onto the fouled membrane.

3.2 Effect of CFV on the hydraulic resistance

Cross flow velocity (CFV) is one of the operating conditions which is a critical parameters that influence the

efficiency of the membrane separation. This is to create an optimal environment for cyclodextrins separation from the reaction mixture as to minimize the fouling effect.

In this study, the investigations of the CFV effect with various permeate flux on the hydraulic resistance during cyclodextrins separation was shown in Figure 3. From this figure, the 1.8 cm/s of cross flow velocity exhibited 5 to 63% more hydraulic resistance compared to 2.6 cm/s and 3.5 cm/s of cross flow velocity. On the other hand, higher cross flow velocity provides less fouling on the membrane surface and pores. This was due to the physical scouring effect at membrane surface. Higher cross flow velocity will drag the back transported particles off the membrane surface and the shear stress generated helps to reduce the rate of cake deposition and the cake thickness [10]. Increasing in hydraulic resistance gave the adverse effect to the permeate flux during CDs separation as shown in Figure 2 which have been discussed before.

Conclusions

The results of these investigations indicated that:

- 1) The cross flow velocity of 1.8 resulted higher value of hydraulic resistance which was 5 to 63% more compared to the cross flow velocity of 2.6 cm/s and 3.5 cm/s.
- 2) The flux declined in ultrafiltration membrane was up to 40% of the initial flux.

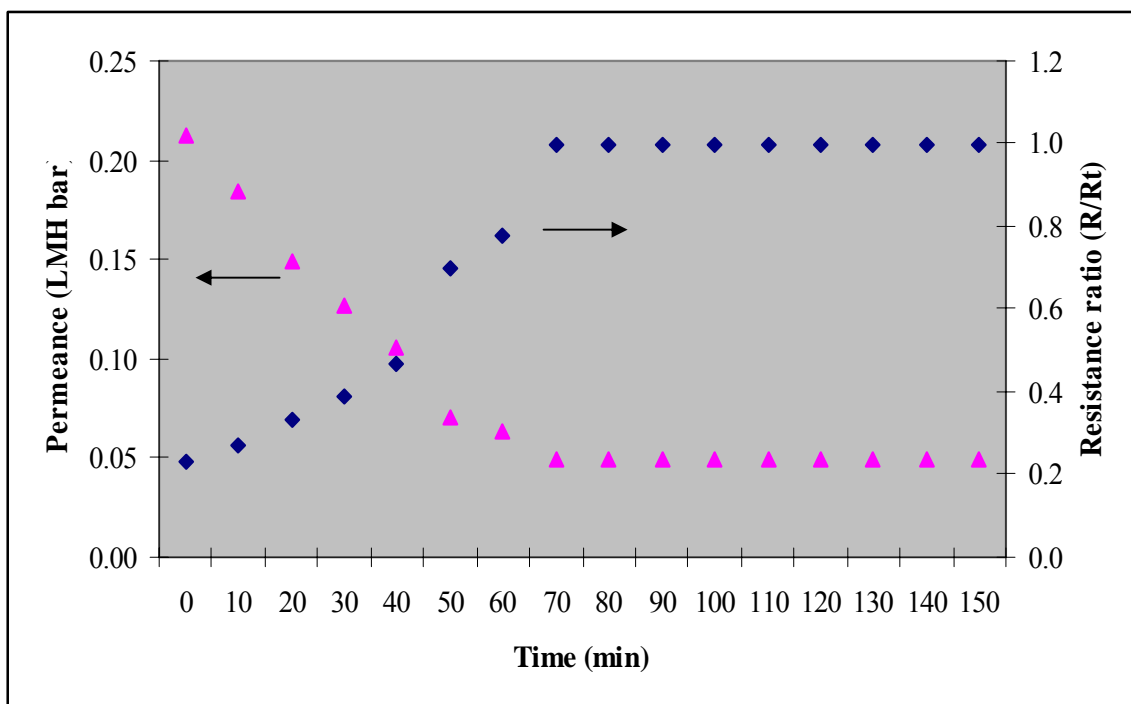


Figure 2: The relation between membrane permeability and resistance ratio in separation of cyclodextrin

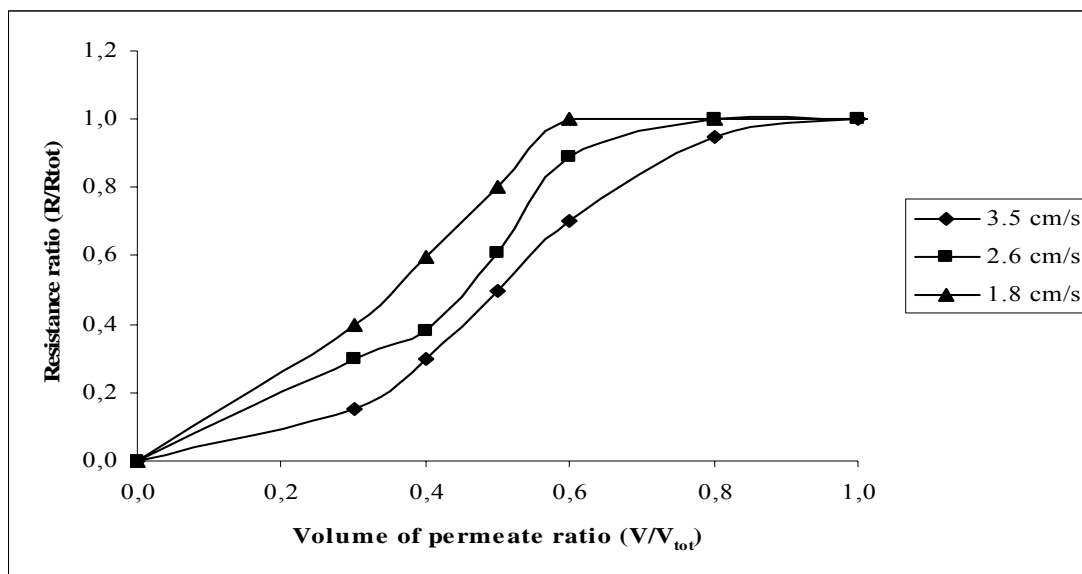


Figure 2: Variation of hydraulic resistance ratio with normalized flux for various CFV

References

- [1] Katchalski-Katzir, E. 1993. Immobilized enzymes learning from past successes and failures. *Trends Biotechnol.* 11:471-478.
- [2] Lopez, C., Mielgo, I., Moreira, M.T., Feijoo, G., Lema, J.M. 2002. Enzymatic membrane reactors for biodegradation of recalcitrant compounds. Application to dye decolourisation. *J. Biotechnol.* 99:249-257.
- [3] Giono, L., Drioli, E. 2000. Biocatalytic membrane reactors: applications and perspectives. *Trends Biotechnol.* 18:339-349.
- [4] Prazeres, D.M.F., Cabral, J.M.S. 2001. Enzymatic membrane reactors. In: Cabral, J.M.S., Mota, M., Tramper, J. (Eds.), *Multiphase Bioreactor Design*. Taylor & Francis, London, pp. 135-180.
- [5] Ibrahim, H. M., Yusoff, W. M. W., Hamid, A. A., Illias, R. M., Hassan, O., Omar, O. 2005. Optimization of medium for the production of β -cyclodextrin glucanotransferase using Central Composite Design (CCD). *Process Biochem.* 40:753-758.
- [6] Biwer, A., Antranikian, G., Heinzle, E. 2002. Enzymatic production of cyclodextrins. *Appl. Microbiol Biotechnol.* 59:609-617.
- [7] Szejtli, J. 1997. Utilization of cyclodextrins in industrial products and processes. *Journal of Material Science.* 7(4): 575-587
- [8] Kim, T.J., Lee, Y.D., and Kim, H.S. 1993. Enzymatic production of cyclodextrins from milled cornstarch in an ultrafiltration membrane bioreactor. *Biotech. And Bioeng.* 41:88-94.
- [9] Cho, J., Amy, G., Pellegrino, J. 2000. Membrane filtration of natural organic matter: factors and mechanisms affecting rejection and flux decline with charged ultrafiltration (UF) membrane. *J. Membrane Science.* 164:89-110.
- [10] Zularisam, A.W, Ismail, A.F., Razman Salim. 2006. Behaviours of natural organic matter (NOM) in membrane filtration for surface water treatment-a review. *Desalination.* 194: 211-232.

Modeling, simulation, and optimization of Continuous Ion Exchange (CIXC) for heavy metal removal from industrial wastewaters

Idral¹ and Saidi, H.^{1*}

¹Membrane Research Unit, Department of Chemical Engineering,
Universiti Teknologi Malaysia, Jln. Semarak 54100, Kuala Lumpur, Malaysia

Tel : +60-3-26154823, Fax : +60-3-26914427

Abstract

Continuous ion exchange system using series plate column for removing of heavy metal ions from industrial wastewaters have been designed and simulated and is now under construction. The pilot plant is expected to remove metal ions according to the DOE regulation and efficient manner and also for minimum generating resin consumption. The concentration of metal ions in the treated wastewater is expected to be the order a few parts per million by volume. Initial investigations are to be carried out using resin Amberlite in a smaller column. Lab-scale operational data will be used for understanding behavior of the system and for on-going scale-up activities using the Matlab Platform.

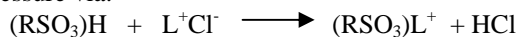
Keywords: continuous ion exchange system, metal removal, series plate column

1. Introduction

There may be a future legislative requirement to remove metal ions from the industry. The worldwide demand for the clean water is projected to increase linearly with the increasing worldwide population and caring of peoples to be the health living. If the both sets of requirements are to be met without significant economic disadvantage for the industrial economy, then new methods for metal ions removal with low or zero metal ions treated must be developed.

The use of the continuous ion exchange system could be potentially meet the needs of a DOE requirements, industrial scale and low cost system. The process outlined below operates continuously; producing a wastewater treated containing below 5 ppm heavy metal ions.

The ion exchange reaction can remove metal ions from the wastewater industrial, room temperature and atmospheric pressure via:



The percentage removal of the systems using continuous ion exchange plate column and resin regeneration has been shown to be comparable to current ion exchange system without plate and batch system. Metal ions concentration will be on the order of a few parts per million since the resin can regenerate again and the resin consumption can be reduced that caused operational cost be less.

Another advantage of the plate column ion exchange system can be applied to the industrial scale and continuous system.

Currently, kinetics data for scale up process for industrial scale are not available in literature. To obtain some information that required for it, a mini-pilot plant is being constructed. The design and Matlab process simulation of this system are presented here.

2. Experimental

2.1. Plate Column ion exchange system

Fig. 1 shows a process flow diagram for the plate column ion exchange system for metal ions removal at the atmospheric pressure. Metal ions removal processing are occurred in the long of column, while resin is fill up onto the plate with the certainty amount with using pulsation system, where the resin can be drawn from the top plate to the plate of below. Partially of the spent resins that going out of the bottom of ions exchange column are feeded directly to the regenerating column to recovery of resin. The recovered resin can be filled to the ion exchange column to the next processes.

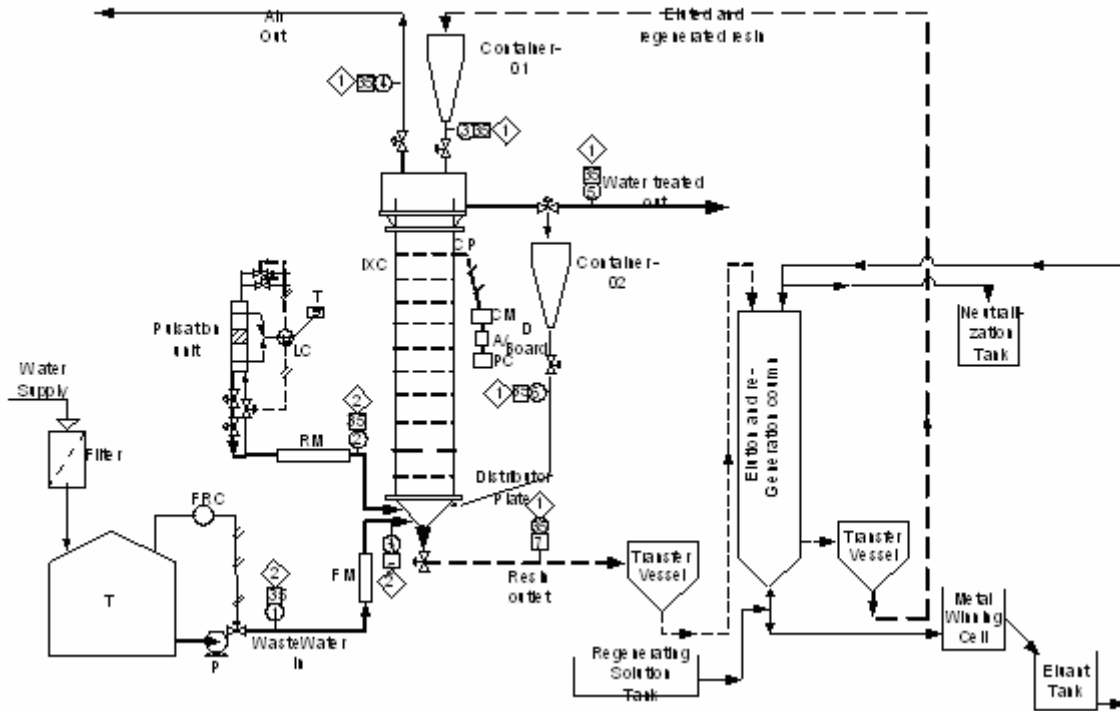


Fig. 1. Process Flow Diagram of the Ion Exchange Plate Column

2.2. Preliminary pilot plant study

Practical operating issues for the ion exchange system have been evaluated using the existing 0.03 m inner diameter glass column. The current process is emulated by various the operating conditions of the existing ion exchange system. This is reached by using the single column for the ion exchange and regenerating column for recovery of the resin. Results of the lab-scale ion exchange system studies are used to fine – tune process simulations for the industrial scale.

2.3. Resin regenerator

Resin is fed to regenerator where metal ions is liberated to recovery of the resin. Air is supplied in regenerator by compressor with the counter current system, where air is entering from the bottom of column, while resin is filling up from top of column. As a result of the resin regenerator is resin regenerated that will be used in the ion exchange column and metal ions solution can release in the bottom of column..

2.4. Rinsing

After the loading and regeneration steps, the resin was rinsed with dematerialized water: ~600 ml will be used for both the resin after loading and for the strong cation resin after the regeneration. After weak cation resin regeneration the rinsing will be continued until the pH of the rinsing metal ions regenerated rose to 2 ± 0.2

3. Modeling procedure

Process simulation using MATLAB has been used to establish the mass and energy balances of interest for equipment selection and design. The governing equation that used to calculate metal ions concentration in solution and resin as respectively is as follow:

$$\frac{\partial^2 C}{\partial z^2} - \frac{U}{De} \frac{\partial C}{\partial z} - \frac{kca}{De} (C - C^*) = \frac{\epsilon}{De} \frac{\partial C}{\partial t} \tag{1}$$

$$De \frac{\partial q}{\partial t} = De \frac{\partial^2 q}{\partial z^2} + \frac{k_l \epsilon_s \cdot \rho_p \cdot q}{\rho_p} (c - c^*) \tag{2}$$

The algorithm of the simulation can be seen in Fig. 2. Resin degradation and percentage removal of metal ions requirements are determined using MATLAB M-file program.

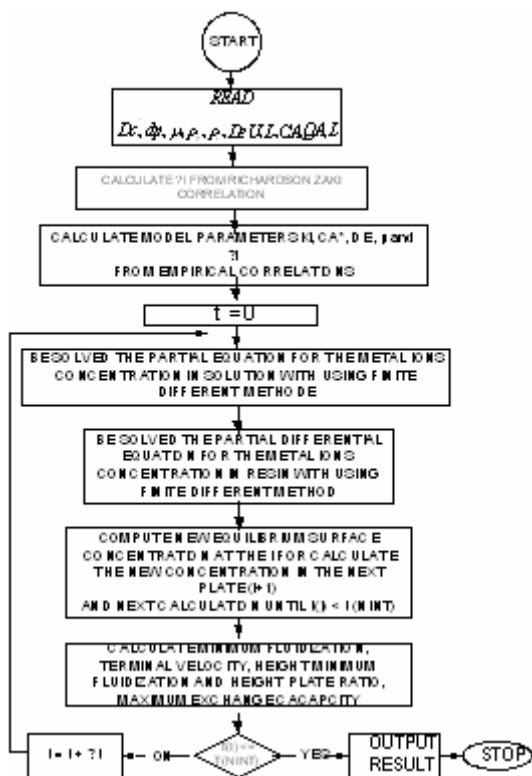


Fig. 2. Algorithm of programming

4. Results and discussion

4.1. Preliminary Lab-scale results

Preliminary investigations using lab-scale ion exchange column indicated that various operating conditions will be gained optimum condition for mass transfer coefficient and effective diffusivities.

The optimum mass transfer and effective diffusivities that obtained, and then can be used to calculate the metal ions concentration a long of column.

Where, based on simulation using hooke jeeves and finite different methods have found mass transfer coefficient is as function of the metal ions concentration in resin that written as $(KCA)=1.864 / (16.836 + QA(i))$ and effective diffusivities $(De) = 1.116E-06 \text{ cm}^2/\text{second}$

4.2. Simulation results

Results of simulation can be explained such figure below:

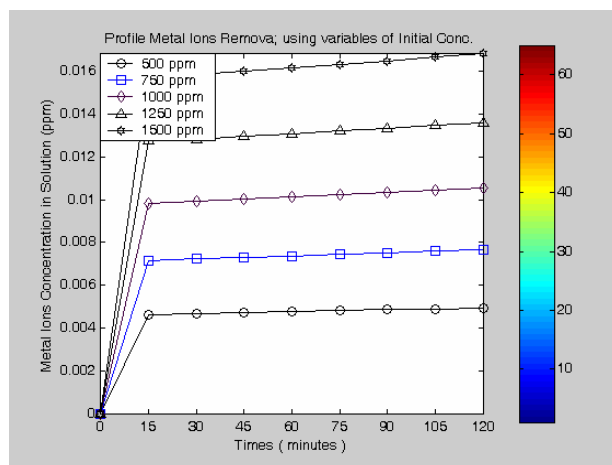


Fig. 3. Profile Metal concentration as the function of Initial Concentration of solution.

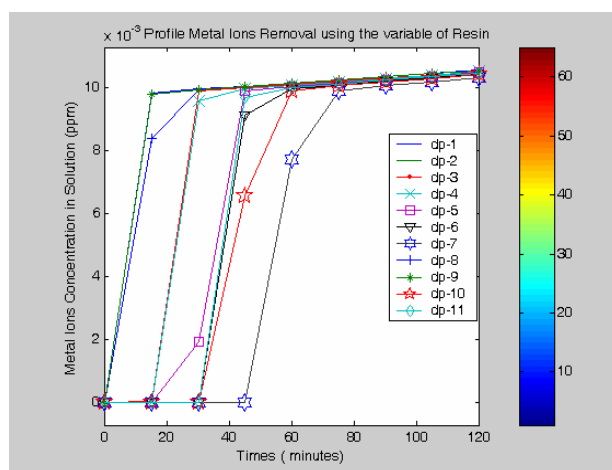


Fig. 4. Profile Metal Ions concentration as the function of variables of resin

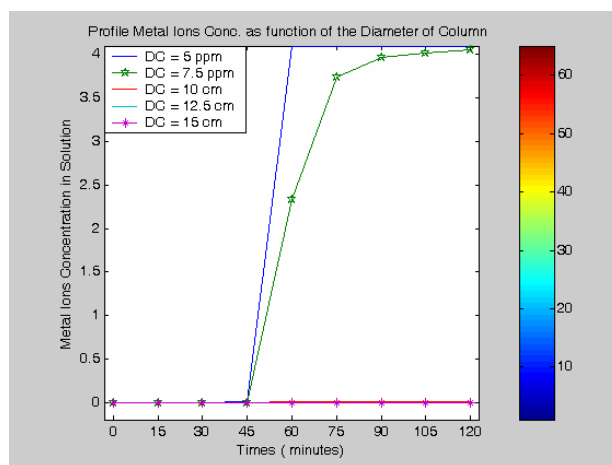


Fig. 5. Profile Metal Ions Conc. as the Function of Diameter of Column

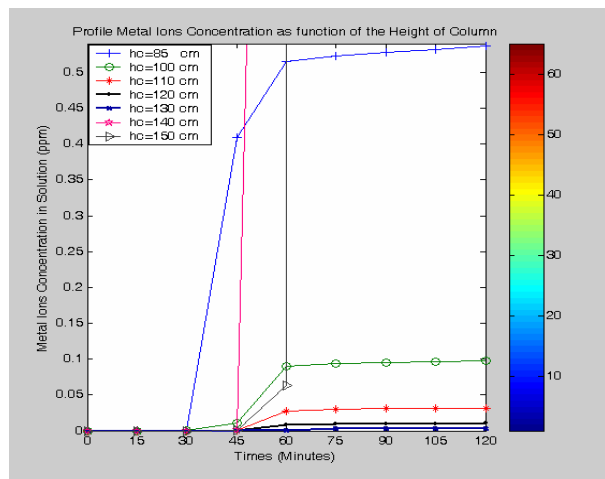


Fig. 6. Profile Metal Ions Conc. as Function of the Height of Column.

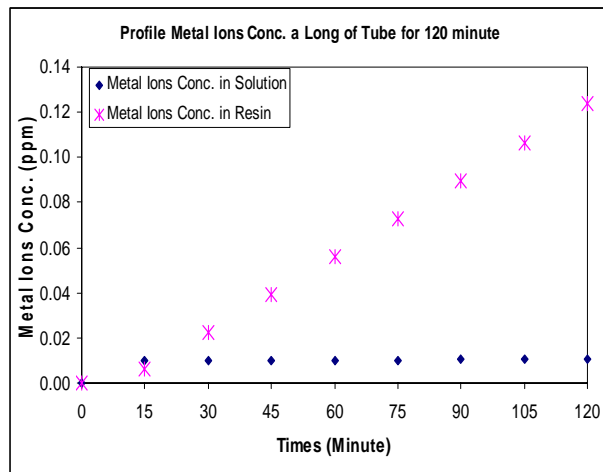


Fig. 9.. Profile of Metal Ions Concentration for 120 minutes processing

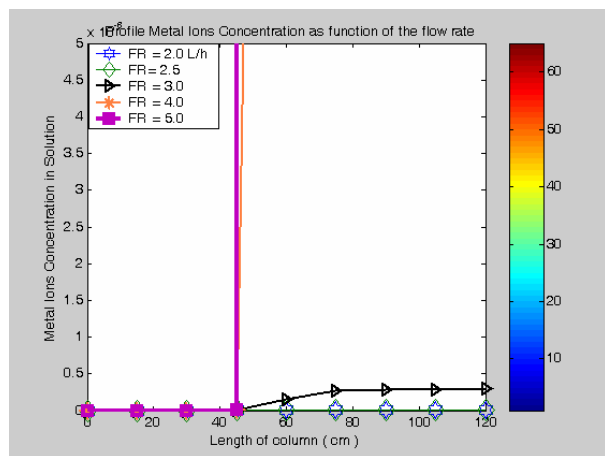


Fig. 7. Profile Metal Ions Conc. as Function of the Flow rate of wastewater.

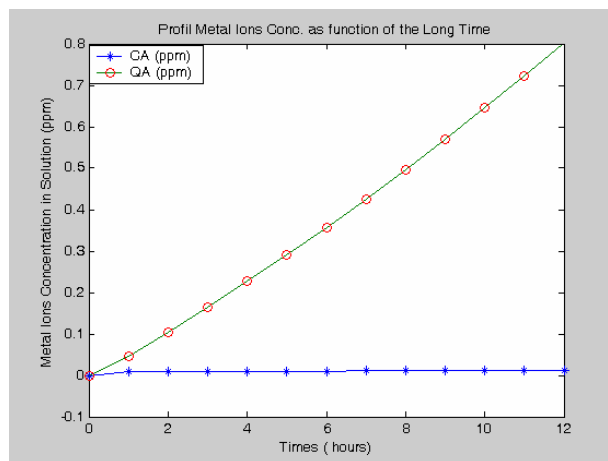


Fig. 10.. Profile Metal Ions Conc. as long as 12 hours predicted by simulation

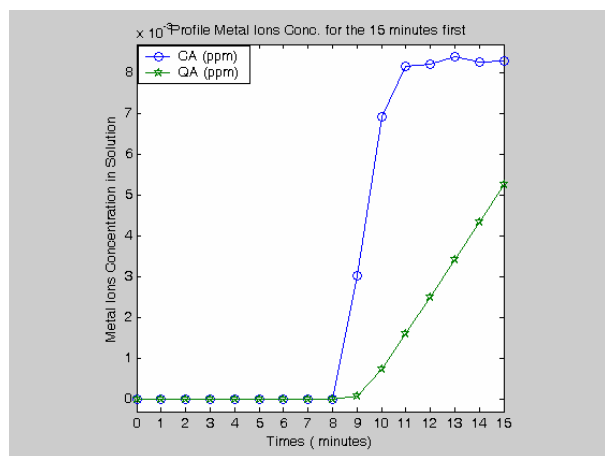


Fig. 8.. Profile Metal Ions Concentration a long of column for 15 minutes first

Base on the results of the simulation, we found optimum operating design of CIXC, with consideration among of variables is as follow;

Amberlite is the best of resin that choose for this system with Grade IRC-718. Where the specification of resin is used to transitional metal chelation, here is shown that the good characteristic of resin have rather a large particle diameter (d_p), less density and the ion exchange capacity is normal range, because until 12 hours operation the percentage removal of metal ions is still in the constrain of the toleration.

The variable of initial metal ions concentration is not be effected to the system, its means that with increasing initial metal ions concentration input, will be automatically be increasing the metal ions concentration in the output, but according by the range variable had selected, actually metal ions concentration output is still a minimum concentration and meet to DoE regulation.

The diameter optimum of column is found at the 10 cm, the results shown that the diameter of column below 10 cm, such that 5 cm and 7.5 cm, given

results of metal ions concentration above of the minimum that required by DoE. While it diameter more than 10 cm, show of results are better, but in this case we choose the optimum diameter is 10 cm, with consideration that : with using diameter more than 10 cm, given results are not significantly different. That means that the selected diameter is cheaper as the economically consideration.. This is phenomenon is met with the trend of equation, Where is increasing the diameter of column will be given contact of reaction is much more, so that possibilities of the ions exchange opportunities are is much more also.

The height variables of column that chosen is 100 cm, 110 cm, 120 cm and 130 cm. Height of column at 130 cm is given maximum percentage removal of metal ions. While the height column at 85 cm, 140 cm and 150 cm are not meet to DoE regulations.

The variable of flow rate is not so effected for the system, since the all variables that selected is can be used and according by the DoE regulation.

5. Conclusions

Bases the all variables and considerations that presented previously, so the optimum design of the system, can be written is as follow :

- Diameter of column, $D_c = 10$ cm
- Height of column, $H_c = 130$ cm
- Resin Amberlite IRC-718, $d_p = 0.5$ mm, density = 1.05 g/ml, tot exchange capacity is 1.0 meq/ml.
- Initial concentration is 1000 ppm of metal ions.
- Flowrate of the system is 5 L/h.

References

- [1] Bolto, B.A., Pawlowski, L.(1987).”Wastewater Treatment by Ion-Exchange.” E. and F.N. Spoon Ltd., London.
- [2] Hubicki, Z., Jakowicz, A., Lodyga, A.(1999).”Application of the ions from waters and sewages. In: Dabrowski, A. (Ed.). Adsorption and Environmental Protection, Studies in Surface Science and Catalysis”. Elsevier, Amsterdam-Lausanne-New York-Oxford-Shannon-Singapore-Tokyo, vol. 120.
- [3] Mehmet Cici, and Yahup Cuci. (1997). “Production of some Coagulant Materials from Galvanizing Workshop Waste”. Vol. 17, No. 7, Waste Management, , pp. 407-410.
- [4] Constantinides, A & Mostoufi, N.(1999).”Numerical Methods for Chemical Engineers with MATLAB Applications.”. Prentice Hall, New Jersey
- [5] Mathews, J.H. & Fink, K.D.(1999).”Numerical Methods Using MATLAB.”Prentice Hall, USA, 3rd Ed.
- [6] Sediawan, W.B. & Prasetya, A.(1997).”Pemodelan Matematis dan Penyelesaian Numeris dalam Teknik Kimia”. 1st Ed. Andi offset. Yogyakarta.
- [7] Yates, J. G. (1983).”Fundamentals of Fluidized-bed Chemical Processes”, .England. Butterworths, Kent, pp. 9-15
- [8] Juang, RS., Lin, SH. And Wang, TY. (2003).”Removal of metal ios from the complexed solutions in fixed bed using a strong-acid ion exchange resin”, Chemosphere, 53, 1221-1228
- [9] Lin, SH., Lai, SL., and Leu, HG. (2000). “Removal of heavy metals from aqueous solution by chelating resin in a multistage adsorption process”, Journal of hazardous Materials, B76, 139-153
- [10] Yu, ping and Luo, Y. (2002).”Novel water treatment process-combined cationic ion-exchange bed and degasifier in a three-phase fluidized bed”, Desalination, 151, 145-152
- [11] Valverde, JL., Lucas, AD, Carmona, M., Gonzales, M. and Rodriguez, JF., (2004).”A generalized model for the measurement of effective diffusion coefficients of heterovalent ions in ion exchangers by the zero-length column method”, Chemical Engineering science, 59, 71-79
- [12] Wang, Y.H, Lin, S.H. and Juang, R.S.(2003).”Removal of heavy metal ions from aqueous solution using various low-cost adsorbents.”Journal of Hazardous Materials B102. 291-302
- [13] Desilva, F.J.(1995).”essential of ion Exchange”. Water Association Technical papers.Resin Tech. Inc.Cherry Hill. NY
- [14] Fernandez, Y., Maranon, E., Cartillon, L. & Vazquez, I.(2005).”Removal of Cd and Zn from inorganic industrial waste leachate by ion exchange”.Journal of hazardous material xxx.p.1-7
- [15] Palacios, V.M., caro, I. & Perez, L.(2001).”Application of ion Exchange Techniques to Industrial Process of metal Ions Removal from Wine”.Adsorption 7. 131-138

- [16] Bektas, N. and Kara, S.(2004). "Removal of Lead from Aqueous Solution by Natural Clinoptilolite: Equilibrium and Kinetics Studies ". Separation and Purification Technology. 39.189-200

Preparation of Polysulfone Ultrafiltration membranes via Microwave Technique

Ani Idris *, Iqbal Ahmad

Department of Bioprocess Engineering
Faculty of Chemical and Natural Resources Engineering
Universiti Teknologi Malaysia, 81310 Skudai, Johor
Tel: +607-5535603, Fax: 607-5581463, Email: ani@fkkksa.utm.my

Abstract

The objective of this paper is to investigate the effect of dope preparation technique on membrane's performance. The novel microwave (MW) technique is employed during the dope preparation and this technique is compared with traditional electro heating (EH) method. Polysulfone (PSf) dope solutions were prepared by dissolution of 20% polymer resin in 80% dimethylacetamide (DMAc) solvent. Asymmetric membranes were casted from these prepared solutions and their performances were then evaluated in terms of pure water permeation, solute permeation rate, PEG separation and molecular weight cut off. The results revealed that dope solutions produced using the microwave technique possess lower viscosity. In addition the membranes produced from the microwave dope solutions exhibits better performance in terms of pure water permeation, solvent permeation rates and solute rejection. The molecular weight cut off at 90% rejection is 1000 kD. The dope preparation time is also reduced thus saves energy and cost.

Keywords: single solvent system, microwave, PSf membrane

1. Introduction

Ultrafiltration is a pressure driven process of separation by membranes, in which solutes are filtered according to molecular size. It has a broad range of application in different fields of biotechnology and chemistry, which has led to the design of large numbers of different types of ultrafiltration membranes. Asymmetric membranes have been prepared by a phase inversion process, which involves the phase separation of a homogenous polymer solution [1].

Currently there is a possibility of enhancing membrane performance beyond the generally recognized intrinsic value for the amorphous polymer. This has been accomplished in a number of different ways for various polymers: polysulfone (PSf), polyethersulfone (PS), polyestercarbonate (PC), polyimide (PI), polyamide (PA) and cellulose acetate (CA) [2, 3]. Among these various membrane materials, polysulfone rank second only to cellulose acetate as a commercial membrane material and are hydrophobic [2]. And this hydrophobic material is commonly used for microporous membranes and contactor applications with an aqueous solution that does not wet the membrane [3].

Polysulfone has repeating phenylene rings, which contribute to high degree of molecular immobility producing, high rigidity, strength, creep resistance and dimensional stability and hence widely used in membrane applications. Many researchers have reviewed the applicability of polysulfone membrane in medical and biotechnological industries.

However, the use of polysulfone for aqueous phase is restricted due to its hydrophobicity and may be improved by modifications of polysulfone. Polysulfone is soluble in a variety of solvents, the most effective of which are the chlorinated aliphatic compounds such as chloroform. In order to avoid the chlorinated chemicals, strong organic solvents such as DMF, NMP, DMAc etc are chosen and since these solvents are not volatile, their removal would be through washing during phase inversion rather than evaporation [4].

The preparation process for dope solution is expensive and time consuming, especially in the case of polysulfone membranes, with minimum dope formulation time of 4 hrs and can sometimes be as long as 24 hrs so as to ensure complete dissolution of polymer. It slows down the development and dissemination of membrane separation technologies in the industry. Historically; dissolution of polymers in dope solution involved mostly electrical heating and physical media-flame. On the other hand the application of microwaves in synthetic organic chemistry has been reported in many publications [5-8]. Many inorganic [9] and organic [10] reactions could proceed under microwave radiation with a much higher rate than conventional methods. The conclusions from the research show that energy transport using microwaves could have a big influence on both the reaction rate and selectivity of polymers[7,8,11,12].

In this study the microwave technique was employed in the preparation of the polymer solution, which consists of polysulfone in DMAc solvent and the performance of these membranes were compared with those prepared using the conventional electro heating method. The performances of the membranes were evaluated using various molecular

* Corresponding Author. E-mail: ani@fkkksa.utm.my,
Tel: +60-7-5535603, Fax: +60-7-5581463

weight polyethylene glycols, PEG. Its molecular weight cut off and flux rates are determined [13].

2. Experimental

2.1. Materials

Polysulfone (Polysulfone-Harz, nominal MW 75000) were used, procured from Organic (Across Organic USA). N,N-dimethyleacetamide DMAc (C_4H_9NO , $M=87.12$ g/mol) were absolute grade (stored under nitrogen atmosphere, in contact with molecular sieves) was purchased from Merck (Merck, Germany) and commercial grade acetone was used without further purification.

2.2. Dope preparation

The dope solutions that consist of 20% PSf and 80% DMAc were prepared by two different methods described below:

i) Microwave method

In this study, Sharp domestic microwave oven model: R-4A53 with the following specifications: rated power output of 850 watts (240V~50 Hz), operation frequency of 2450 MHz is used. A 500 ml Schott Duran is used as the sample reaction vessel at atmospheric pressure. Mercury thermometer was used manually to control the temperature at every 20 sec. The temperature of the dope solution was kept at 85-95 °C. Heating time by microwave was 10 minutes.

ii) Traditional electro heating (EH) method

In the traditional electro heating, the polymer dissolution process is carried out in a 1 liter 3 necked round bottomed flask with stirrer and condensers as described elsewhere [14] The electro thermal heater used is (230V~50/60 Hz, 300 Watts) under the brand name Barnstead/electro thermal. The dope temperature is kept constant at 90 °C for 6 hrs.

2.3. Determination of permeation flux and solutes rejection

The performance of the membranes prepared using these two techniques were evaluated in terms of pure water permeation fluxes (J_w), solvent permeation fluxes (J_s) and rejection rate (SR) in a test cell described elsewhere [15]. A minimum of three flat sheet samples was prepared for each technique and the average data are tabulated. Pure water permeation fluxes (J_w) and solutes water permeation fluxes (J_s) of membranes were obtained as follows:

$$J = \frac{Q}{\Delta T \times A} \quad (1)$$

where J is the permeation flux of membrane for PEG solution ($L m^{-2} h^{-1}$) or pure water and Q is the volumetric flow rate of permeate solution.

Solute rejection of membranes were evaluated with various molecular weight PEG solutions ranging from 600

to 35000 kDa at 10 bar. The concentration PEG solution used is 1000ppm. The concentration of the feed and permeate solution were determined by the method described as follows:

Reagent A: 5% (w/v) $BaCl_2$ in 1 N HCl (100 ml)

Reagent B: 2% (w/v) KI diluted 10 times + 1.27g I_2 .

Four milliliters of sample solution was added to 1 ml reagent A. To this mixture 1 ml of reagent B was added. Colour was allowed to develop for 15 min at room temperature, and adsorption was read using a spectrophotometer at 535 nm against a reagent blank [16]. The membrane rejection (R) is defined as

$$R = 1 - \frac{C_p}{C_f} \quad (2)$$

where C_f and C_p are the polyethylene glycol concentration in the feed solution and permeate solution, respectively. The concentration of PEG was determined based on absorbency in a UV-spectrophotometer at a wavelength of 535 nm.

3. Results and Discussion

3.1 Dope Solution Properties

The time taken for preparing the dope solutions using the microwave technique (MW) is very short, approximately 30 minutes (0.5hr) compared to the traditional electro thermal heating (EH) method as can be observed in Table 1. The time reduction of 70% is because of the efficient heating of the viscous mass inside the reactor, by the penetrating microwaves. Volumetric heating by microwave is not limited by slow heat conduction from the reactor surface to the inside, as occurs with conventional heating. In addition the dope solution obtained using the MW technique is much lower resulting in ease in casting and spinning.

Table 1: Dope solution properties obtained by microwave and traditional electro heating method

Dope preparation techniques	Dope composition (wt %)	Preparation time (hrs)	Viscosity at 28 °C (cps)
Microwave method	PES/DMAc 20/80	0.5 at 90-95 °C	518
Traditional EH method	PES/DMAc 20/80	6 at 90-95 °C	568

3.2. Performance of the membranes

In addition, the membranes produced using microwave method gives much higher pure water permeation (J_{pw}) and permeate flux (J_s) rates than those produced by electro thermal heating method. This can be clearly observed from the graphs plotted in Figure 1. The pure water permeation

and permeation flux for these membranes were approximately doubled which means increase in productivity.

The same phenomenon is observed for their rejection rates as depicted in Figure 2. Microwaved prepared membranes exhibit better rejection rates. The molecular weight cut offs (MWCO) at 90% rejection rate for membranes prepared using the microwave techniques are much lower than those prepared using the conventional electro heating method.

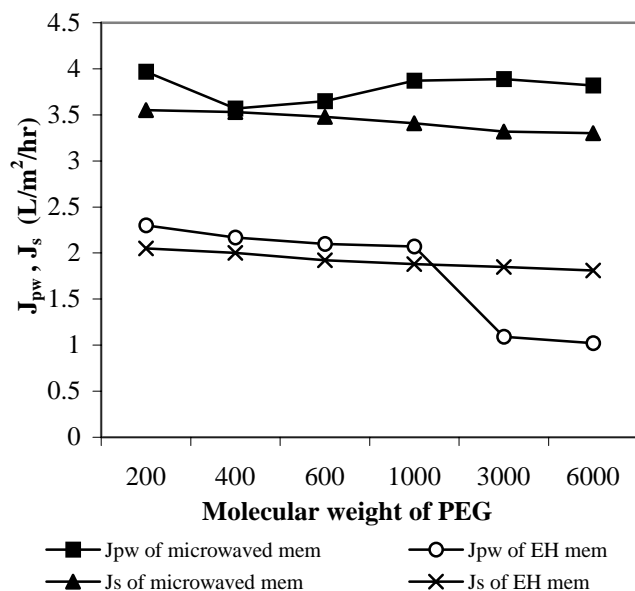


Figure 1 Pure water permeation flux (J_{pw}), permeate rates (J_s) versus molecular weight s of PEG using the two techniques

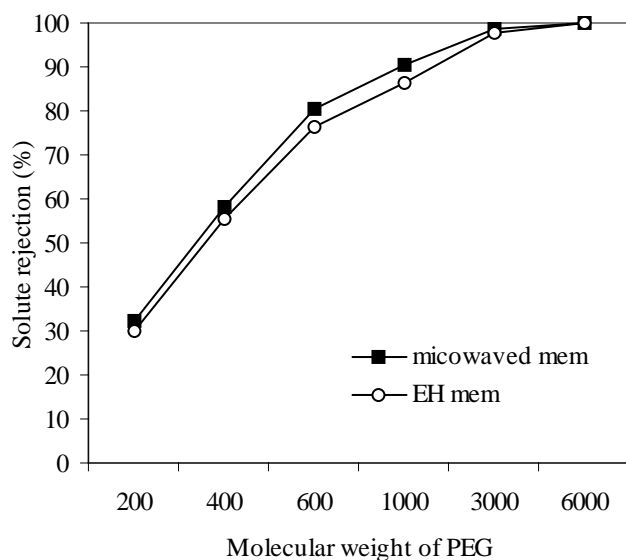


Figure 2 PEG rejection rate versus molecular weights of PEG using the two techniques

Microwave heating has some other advantages. The membranes synthesized by microwave heating are stable

and dense, and their thickness is well controlled due the lower viscosity dope produced.

During the phase inversion process the low viscosity dope solution produced using microwave technique probably results in its rapid diffusion compared to the coagulant (water) inflow during the wet phase inversion process thus offering macrovoids free porous inter-connecting channels generating the formation of smaller and closer cells with a thin skin layer. The formation of such thin-skinned membranes produced probably explained for the high solute separation as well as the high flux exhibited.

4. Conclusion

In summary membranes produced from dope solutions prepared using the microwave technique are far more superior in terms of flux rate compared to those prepared using the conventional electro heating method. The rejection rate of the membranes is higher. Apparently microwaves result in rapid heating which accelerates the reaction rate and in this case the dissolution of the polymer in the solvent. Thus this novel microwave method has proved to be a very rapid and economical process and has the potential to bring membrane technology to new heights.

Acknowledgements

Financial support from the Ministry of Science, Technology and Environment through the IRPA funding vote no 74246 and 74170 is gratefully acknowledged.

References

- [1] Reddy, A. V. R., Mohan, D. J., Bhattacharya, A., Shah, V. J., and Ghosh, P. K. (2003). Surface modification of ultrafiltration membranes by preadsorption of a negatively charged polymer: I. Permeation of water soluble polymers and inorganic salt solutions and fouling resistance properties. *Journal of Membrane Science*. 214(2): 165-350
- [2] Sivakumar, M., Mohan, D. R., and Rangarajan, R. 2006. Studies on cellulose acetate-polysulfone ultrafiltration membranes: II. Effect of additive concentration. *Journal of Membrane Science* 268(2): 208-219
- [3] Cheng, Kwok-Shun; Gates, T. Dean; Yen, Larry Y. and Patel, Rajnikant B. 2000. Skinned hollow fiber membrane and method of manufacture. *United States Patent* 6921482
- [4] Morrison, R. T., Boyd, R. N. (1978). *Organic chemistry 3rd ed.* chapter 1 pp 30-32
- [5] Sajitha, C.J., and Mohan, D. 2003. Studies on cellulose acetate-carboxylated polysulfone blend ultrafiltration membranes - Part II. *Polym Internatioinal* 52(1):138-145.
- [6] Caddick, D. 1995. Microwave assisted organic reactions *Tetrahedron*. 51(38): 10403-10650
- [7] Abramovitch, R. A. 1991. For relevant papers and reviews on microwave assisted chemical reactions. See; *Organic Preparation Process International*. (23), 683

- [8] Perreux, L., and Loupy, A. 2001. A tentative rationalization of microwave effects in organic synthesis according to the reaction medium, and mechanistic considerations *Tetrahedron*. 57(5): 9199.
- [9] Komarneni, S., Roy, R., and Li, Q. H.(1992). Microwave-hydrothermal synthesis of ceramic powders. *Mater. Res. Bull.* 27:1393.
- [10] Baghurst, D. R and Mingos, D. M. P. J. (1992). Microwave chemistry in liquid media *chem. Soc., Chem. Commun.* **21**: 674-677
- [11] Kingston, H., Haswell, S. 1997. Microwave-Enhanced Chemistry. *ACS*, 1997
- [12] Deshayes, S., Liagre, M., Loupy, A., Luche, J. L., and Petit, A. 1999. Microwave activation in phase transfer catalysis *Tetrahedron*. 55(36): 10851-10870.
- [13] Schwarz, H. H., Bossin, E. and Fanter, D. 1982. Skin layer characterization of anisotropic membranes for ultrafiltration *Journal of Membrane Science*. 12(1): 101-106
- [14] A. Idris, Fabrication and Optimisation of Asymmetric Hollow Fiber membranes for reverse osmosis, PhD dissertation, UTM. 2002
- [15] Ani Idris, Farida Kormin, Zakiamani, (2005). Curing Temperature Effect on the Performance of Thin Film Composite Membrane *Jurnal Teknologi F* **43** 51-64
- [16] Ashutosh D. Sabde, M.K.Trivedi, V. Ramachandhran, M.S. Hanra. 1997. Casting and characterization of cellulose acetate butyrate based UF membranes, *Desalination* 114:

Influence of bore fluid flow rate to ultrafiltration hollow fiber membrane performance

Ani Idris*, Sze- Yean Lee

Department of Bioprocess Engineering, Faculty of Chemical and Natural Resources Engineering
Universiti Teknologi Malaysia 81300 UTM Skudai, Johor Bahru.
Tel: +607-553560, Fax: +607-5581463 *ani@fkkksa.utm.my

Abstract

Hollow fibers for ultrafiltration (UF) have been fabricated from a dope solution containing Polyethersulfone (PES)/ *N*-methyl- 2- pyrrolidone (NMP)/ Polyvinyl-pyrrolidone (PVP) with a weight ratio of 20/75/5. The dry jet wet spinning process that introduces an evaporation step before the immersion in a non-solvent bath was chosen to fabricate defect free asymmetric hollow fiber membranes. Effects of bore fluid chemistry on morphology and performance of membranes have been investigated. Bore fluid flow rate (BFR) was varied from 1.0 to 3.5 ml/min with intervals of 0.5 ml/ min. By maintaining the ratio of BFR to polymer extrusion rate at 1/3, the results showed that the wall thickness of the hollow fiber increased as the BFR increased. The pure water permeation (PWP) rate is expected to increase linearly with the wall thickness to an optimum value and then decrease as the membrane wall continues to thicken. The performance of the UF hollow fiber membranes were evaluated in terms rejection rate and permeate flux using polyethyleneglycol (PEG) 35000 as the feed solution. The scanning electron microscope (SEM) is used to confirm the changes in morphology. In this study, the other spinning parameters such as polymer concentration, air gap distance and take up velocity is kept constant.

1. Introduction

Ultrafiltration (UF) membrane is widely used for concentration and clarification of a variety of food and biological systems. Since its invention, many researchers have tried to extend and improve the newly developed technologies to obtain high performance asymmetric hollow fibers.

One important goal in membrane technology is to control membrane structure and thus membrane performance. This objective is not easy to achieve because membrane structure and performance depend upon various factors such as choice of polymer, solvent and non-solvent, composition and temperature of coagulant and casting solution among others. Furthermore, by changing one or more of these variables, which are dependent on each other, membrane structure may be affected quite significantly.

Fane [1] studied the properties of UF membranes which make them susceptible to fouling and also described various types of flux decline. Tietze et al.[2] reported determination of pore size distribution based on flux and pressure measurements with immiscible liquids. Van't Hof [3] and Li et al. [4] invented the dual-bath technique to fabricate PES asymmetric hollow fibers with skin thickness of 200-300 nm for gas separation using a ternary spinning dope.

Chung et al. [5] fabricated PES hollow fibers with a skin layer of 420- 470 Å for the first time by using a binary

fluid on the performance of final fibers, and reported that bore fluid flow rate would have important influence on its performance.

In this paper, a unique dope mixture consisting of PES and PVP is dissolved in NMP and the influence of different bore fluid flow rates towards the membrane performance, in terms of its pure water permeation rate and solute rejection rate for PEG 35000 is investigated.

2. Experimental

2.1. Fabrication of hollow fiber membrane

Dope solution consisting of 20% Polyethersulfone (PES), 75% *N*-methyl-2-pyrrolidone (NMP) and 5% polyvinylpyrrolidone (PVP K30) is used in this study. The dope solution was prepared by dissolving predetermined amounts of PES and PVP K30 in NMP and was debubbled before being used.

The PES hollow fiber membranes are spun by the forced convective dry/wet spinning method described elsewhere [6]. A metering gear pump pumped the dope to the spinneret. The polymer solution flowed through a ring nozzle while the bore fluid was injected through the inner tube of the spinneret.

The fiber passed through a 10cm air gap (filled with Nitrogen gas) before it entered the coagulation bath. Water was used as the external coagulant to yield an outer selective dense layer. After spinning and coagulation, the fiber was rinsed with water and post treated with sodium hypochlorite and glycerol to remove excess PVP and impart the rewetting and flux characteristics of the fibers. The fibers were then dried in air on tissue papers at room temperature.

Corresponding Author. E-mail: ani@fkkksa.utm.my,
Tel:+607-5535603, Fax:607-5541463

dope (PES in NMP) in a wet spinning process. They investigated the effect of solvent concentration in the bore

2.2. Scanning Electron Microscope

For morphology study, the spun fiber were immersed in liquid nitrogen and fractured to obtain tidy fiber cross-sections. These samples were then attached to a carbon holder and sputtered with gold to prevent charging up of the surface by the electron beam. A narrow beam of electrons with kinetics energies in order of 1 – 25kV hits the membrane sample, and low- energy electrons were liberated from the atoms in the surface to create the image on the micrograph. Cross sections of the hollow fiber membranes images were obtained using the SUPRA 35VP FE- SEM.

2.3. Measurement of rejection rate of hollow fibers

Hollow fibers were tested in bundles of 30 fibers, each in the length between 22 - 23cm. The fibres were potted with epoxy resin and mounted in a test module described elsewhere [6]. The operating pressure and feed flow- rate were 40 psig and 1.5L/min, respectively. The feed solution was supplied to the shell side of the hollow fibers and permeate was collected from the bore side. The PEG solute concentration used was 200 ppm. The concentration of the feed and permeate solution were determined by the method described as follows:

Reagent A: 5% (w/v) BaCl₂ in 1 N HCl (100 ml)

Reagent B: 2% (w/v) KI diluted 10 times + 1.27g I₂.

Four milliliters of sample solution was added to 1 ml reagent A. To this mixture 1 ml of reagent B was added. Colour was allowed to develop for 15 min at room temperature, and adsorption was read using a spectrophotometer at 535 nm against a reagent blank [7].

The rejection rate, f , was calculated by the following equation:

$$f(\%) = \frac{C_{feed} - C_{permeate}}{C_{feed}} \times 100 \quad (1)$$

3. Results and discussion

3.1. Pure Water Permeation Rate

The pure water permeation rate of the ultrafiltration membrane was very low, less than 0.15 L/m².hr, when the bore fluid flow rate was 1 ml/ min as can be observed from Figure 1. As the BFR increases to 1.5 ml/min, the PWP rate increases to 0.75 L/m².hr. This phenomena could be due to the outer skin formed was dense enough to produce high resistance to water permeation.

The PWP decreases slowly from 0.75 L/m².hr to 0.47 L/m².hr as the bore fluid flow rate increases from 1.5 ml/min to 3 ml/min. This could be due to the increase in membrane's skin thickness, which then resulted in the decrease of permeation rate. As it reaches a certain critical point above BFR of 3 ml/min and above, the PWP

accelerates to 1.3 L/m².hr, due to the formation of a more porous outer skin

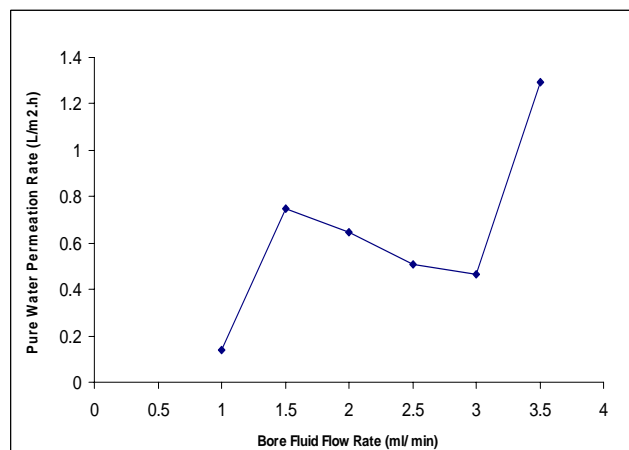


Figure 1 Variation of PWP rate with different bore fluid flow rate (ml/ min)

3.2. Solute Rejection

Figure 2 shows that the solute rejection rate increases as the bore fluid injection rate increases from 1.5 ml/ min to 2.5 ml/ min. The membrane spun with low bore fluid flow rate (1 ml/ min) tend to form a layer of dense skin, which offered an excellent rejection rate up to 99.7% but low PWP rates. As the BFR slowly to 1.5 ml/ min, rejection rate decreases to 92%.

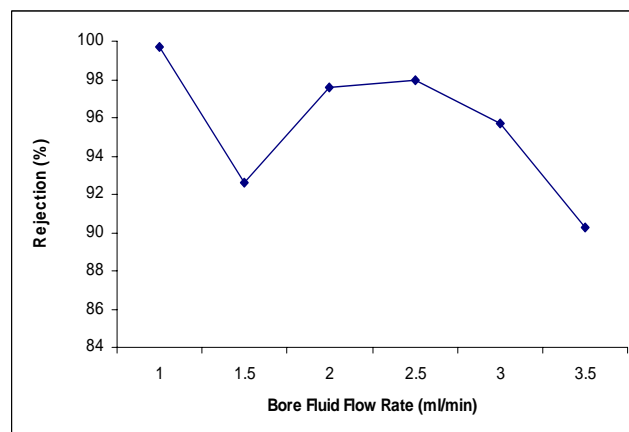


Figure 2 Variation of rejection rate (%) with different bore fluid flow rate (ml/ min)

Further increase of BFR to 2.5 ml/ min reveals an increment in rejection rate from 92% to 97%. This result is consistent with the observed PWP rate, whereby increase in rejection rate is linked to flux decrease. When the bore fluid injection rate is further increased to 3.5 ml/ min, the skin layer formed becomes even more porous, thus producing lower rejection rates. In order to keep the thickness of the membrane constant the ratio of BFR to DER is kept at 1:3. Thus as the BFR increases, the DER also increases, in fact 3 times more. Hollow fibers spun at such high extrusion rates has the tendency to tear the molecular chains apart

thus generating slight imperfections in skin layer, resulting in rejection rates to decline [8].

3.3. Electron micrograph observation

Figure 3 shows the electron micrograph picture of the cross sections of hollow fiber membranes at different bore fluid flow rate (BFR): (a) BFR 1.0 (b) BFR 1.5 (c) BFR 2.0 (d) BFR 2.5 (e) BFR 3.0 (f) BFR 3.5. These pictures

1. The pure water permeation rate increase as the bore fluid flow rate increase to a certain critical point before it begins to decrease probably due to change in molecular orientation in membrane's skin layer.
2. The rejection rate of membrane remains high for all bore fluid flow rates ($> 90\%$), with slight decrement at high rejection rates (> 2.5 ml/min) due to the porous structure formed in membrane.

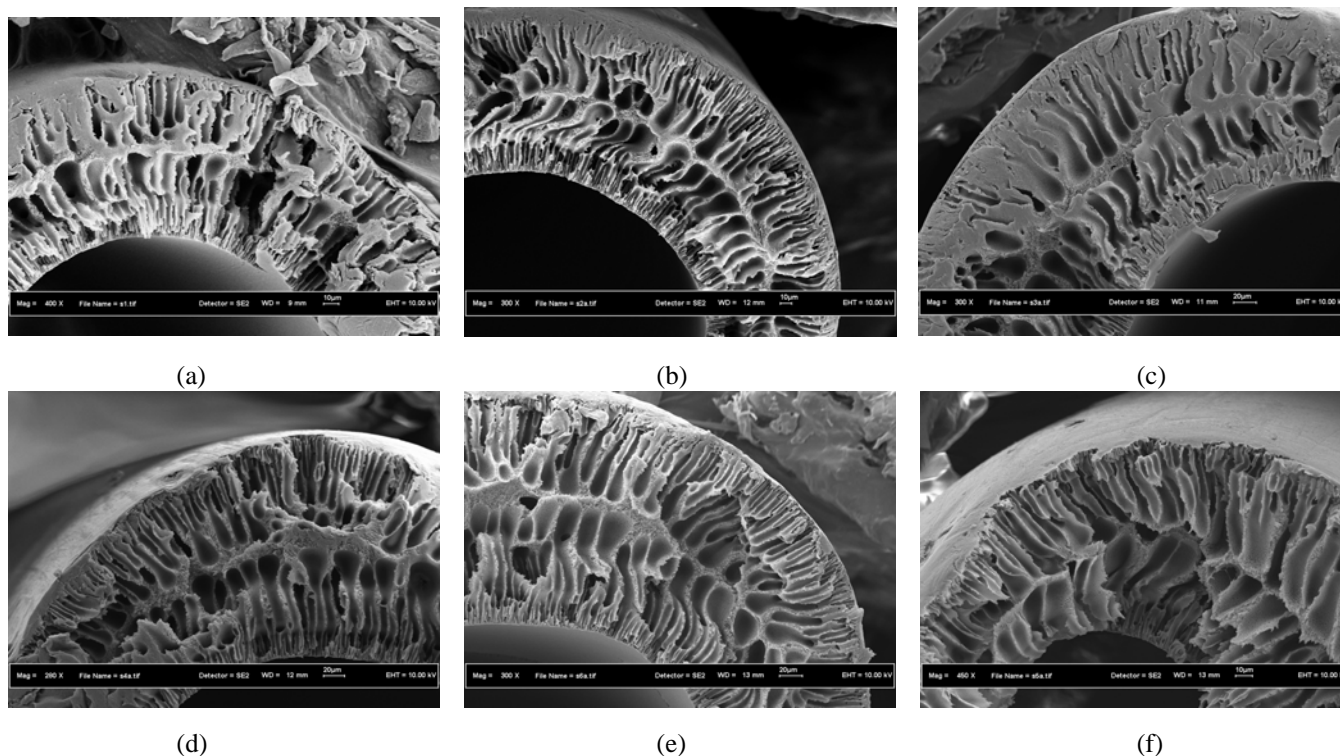


Figure 3 The cross- section of hollow fiber membranes at different bore fluid flow rate: (a) BFR 1.0 (b) BFR 1.5 (c) BFR 2.0 (d) BFR 2.5 (e) BFR 3.0 (f) BFR 3.5 ml/min

indicate the presence of two dense layers both on the bore side and shell side of the hollow fibers. It can be seen that there is a boundary between the region of small finger-like voids and that of larger finger-like voids near the shell side surface of the hollow fibers [9]. This is considered to be the boundary where the flows of the gelatine media by diffusion from the bore side and from the shell side meet. Very thin asymmetric skin layer were observed for membrane (a), (c) and (d) and the layer of finger like voids on the shell side is thinner than the layer of finger-like voids near the bore side surface of the hollow fibers which probably explained for the high rejection rate.

4. Conclusion

From the above experimental results and theoretical considerations, the following conclusions can be drawn for the ultrafiltration performance of hollow fiber membranes:

Acknowledgements

Financial support from the Ministry of Science, Technology and Environment through the IRPA funding vote no 74246 is gratefully acknowledged.

References

- [1] A. G. Fane and C. J. D. Fell. 1987. A review of fouling and fouling control in ultrafiltration. *Desalination*, 62: 117 - 136.
- [2] R. Tietze, P. Löttsch and K. Franke, in: *Synthetic Polymeric Membranes*, Water de Gruyter, Berlin, New York, 1987.
- [3] J.A. van't Hof. 1988. Wet spinning of symmetric hollow fiber membranes for gas separation, Ph.D. Thesis, University of Twente, the Netherlands.
- [4] S.G. Li, G.H. Koops, M.H.V. Mulder, T. van den Boomgard, C.A. Smolders. 1994. Wet-spinning of integrally skinned hollow fiber membranes by a modified

dual-bath coagulation method using a triple orifice spinneret, *J. Membr. Sci.* 94: 329.

[5] T.S. Chung, S.K. Teoh, X. Hu. 1997. Formation of ultrathin highperformance hollow fiber membranes, *J. Membr. Sci.* 133: 161.

[6] A. Idris. 2002. Fabrication and Optimisation of Asymmetric Hollow Fiber membranes for reverse osmosis, PhD dissertation, Dept of Bioprocess Engineering, UTM.

[7] Ashutosh D. Sabde, M.K.Trivedi, V. Ramachandhran, M.S. Hanra. 1997. Casting and characterization of

cellulose acetate butyrate based UF membranes, *Desalination* 114: 223- 232.

[8] A.F. Ismail, M.I.Mustaffar, R.M. Illias, M.S. Abdullah. 2006. Effect of dope extrusion rate on morphology and performance of hollow fiber membrane for untrafiltration, *Separation Purification Technology* 49:10- 19.

[9] S.Mok, D.J. Worsfold, A.E. Fouda, T. Matsuura, S.Wand, K. Chan.1995. Study on the effect of spinning condition and surface treatment on the geometry and performance of polymeric hollow- fiber membranes, *J. Membr. Sci.* 100:183- 192.

NN with DTW-FF Coefficients and Pitch Feature for Speaker Recognition

Rubita Sudirman¹, Sh-Hussain Salleh¹, Shaharuddin Salleh²

¹Faculty of Electrical Engineering

²Faculty of Science

Universiti Teknologi Malaysia, 81310 Skudai, Johor

Tel.: 607-553 5738, Fax:607-553 5681, rubita@fke.utm.my

Abstract

This paper proposes a new method to extract speech features in a warping path using dynamic programming (DP). The new method presented in this paper described how the LPC feature is extracted and those coefficients are normalized against the template pattern according to the selected average number of frames over the samples collected. The idea behind this method is due to neural network (NN) limitation where a fixed amount of input nodes is needed for every input class especially in the application of multiple inputs. The new feature processing used the modified version of traditional DTW called as DTW-FF algorithm to fix the input size so that the source and template frames have equal number of frames. Then the DTW-FF coefficients are retained and later being used as inputs into the MLP neural network training and testing. Thus, the main objective of this research is to find an alternative method to reduce the amount of computation and complexity in a neural network for speaker recognition which can be done by reducing the number of inputs into the network by using warping process, so the local distance scores of the warping path will be utilized instead of the global distance scores. The speaker recognition is performed using the back-propagation neural network (BPNN) algorithm to enhance the recognition performance. The results compare DTW using LPC coefficients to BPNN with DTW-FF coefficients; BPNN with DTW-FF coefficients shows a higher recognition rate than DTW with LPC coefficients. The last task is to introduce another input feature into the neural network, namely pitch. The result for BPNN with DTW-FF plus pitch feature achieved its high recognition rate faster than the combination of BPNN and DTW-FF feature only.

Keywords:

dynamic time warping, normalization, linear predictive coding, pitch feature, back-propagation neural network

1. Introduction

Speech recognition describes a group of special technologies that allow callers to speak words, phrases, or utterances that are used to control some particular applications. In the case of voice processing, speech recognition is used to replace touch-tone input, make for more intuitive menu structures, and add a level of simplicity and security to some systems. Speech recognition, on the other hand, is a technology that uses the spoken word as input that has an effect on the logic flow and execution of the program in query.

The recognition technique like the NN has been widely used as a recognition engine in speech recognition and other pattern recognition applications. There are also various form of input can be recognized by the network depending on the network setting whether it can accept and process single input or multiple inputs at a time. The mechanism of the network itself plays an important role to determine the suitable parameters and inputs for a particular application. In speech recognition application, a back-propagation NN can be used as the recognition engine and modified according to the norm of the problem.

From the literature reviews, past and most current research are using the global distance scores [2][3][4][9], or LPC coefficient as an input to the neural network one sample at a time. In that respect, a new method called Dynamic Time Warping Frame Fixing (DTW-FF) which is based on DP [1] is proposed to extract another form of feature which has a smaller number of input size so that it can reduce the amount of computation and network complexities in the back-propagation neural network.

Neural Network is chosen as the back-end recognition engine due to its past good and reliable performances in speech recognition. As mentioned in the earlier paragraphs, NN is considered as one of the popular method used especially when dealing with isolated word speech recognition. This study considers mainly on isolated words, NN is chosen as an engine to perform the recognition task. Since the main task of the study is to find an alternative way of reducing the number of inputs into the NN, this should light up a new form of input representation into the NN, which is simpler and smaller when compared to LPC feature.

2. Objective

The main objective of the study is to propose a method of time normalization to speech signals so that it can be used concurrently with other samples at a time using the back-propagation NN. Result from the time normalization which is composed from traditional DTW which then we called as the DTW-FF is used with another feature, namely pitch to compare the recognition performance using DTW-FF itself and both features.

3. Methods of Feature Extraction

The features used in this study are extracted separately before they are combined for the input into the NN. They are extracted as follows:

Step 1: LPC Feature Extraction

The LPC feature is extracted from the raw signal and it goes through some procedures which involve pre-emphasis, frame blocking, windowing, autocorrelation, and LP coefficient computation itself. The details can be found in [4][5][8].

Step 2: DTW-FF Feature Extraction

Every frame in a template and test speech pattern must be used in the matching path. Considering DTW path type 1 (Fig. 1), if a point (i, j) is taken, in which i refers to the test pattern axis (x-axis), while j refers to the template pattern axis (y-axis), a new path must continue from previous point with a lowest distance path, which is from point $(i-1, j-1)$, $(i-1, j)$, or $(i, j-1)$. If a reference template with feature vector R and an input pattern with feature vector T , each has of N_T and N_R frames, the DTW is able to find a function $j=w(i)$, which maps the time axis i of T with the time axis j of R . The search is done frame by frame through T to find the best frame in R , by making comparison of their distances.

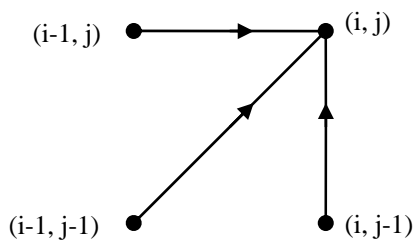


Fig. 1 DTW path type I

The global distance, D of a warping path is defined as

$$D(i, j) = \min[D(i-1, j-1), D(i-1, j), D(j-1, i)] + d(i, j) \quad (1)$$

The DTW-FF feature is extracted using the DTW-FF algorithm which is designed based on the traditional DTW. The mechanism in the DTW algorithm is meant to calculate and find the shortest warping path between an input and the reference template. Similar occurs in the DTW-FF

algorithm except that the algorithm is altered in such a way that it can fix the input frames number to the same amount as the frame number of the reference template. This is what we called as DTW frame fixing or DTW-FF in short. This frame alignment is also known as the expansion and compression method [6].

In this algorithm, frame fixing mechanism also includes the compression and expansion of frame numbers of the input, the reference template frame number is fixed, thus the input is adjusted according to the reference frame numbers. If the input has less frame number, it means that the frame number has to be expanded otherwise if the input frame number is more than the reference frame number, then the input frame number has to be compressed. The expansion and compression has to follow some rules according to the slopes of the warping path:

i- Slope is ~ 0 (horizontal line)

When the warping path moves horizontally, the frames of the speech signal are compressed. The compression is done by taking the minimum calculated local distance amongst the distance set, i.e. compare $w(i)$ with $w(i-1)$, $w(i+1)$ and so on, and choose the frame with minimum local distance. The search is represented as

$$F^- = F(\min\{d_{(i,j)...(i,j)}\}) \quad (2)$$

ii- Slope is $\sim \infty$ (vertical line)

When the warping path moves vertically, the frame of the speech signal is expanded. This time the reference frame gets the identical frame as $w(i)$ of the unknown input source. In other words, the reference frame duplicates the local distance of that particular vertical warping frame. The expansion can be represented as

$$F^+ = F(r(i)) \quad (3)$$

iii- Slope is ~ 1 (diagonal)

When the warping path moves diagonally from one frame to the next, the frame is left as it is because it already has the least local distance compared to other movements.

The distance is calculated using Euclidean distance measure. For a set of LPC coefficients with p feature vectors, which is from $j=1, 2, \dots, p$ of (x, y) coordinate, the distance is calculated as

$$d(x, y) = \sqrt{\sum_{j=1}^p (x_j - y_j)^2} \quad (4)$$

Step 3: Pitch Feature Extraction and Optimization

Pitch is another feature considered in this study to ace the recognition rate of using the DTW-FF coefficients only. An application that strictly requires pitch into the system is the cochlear implant; the implant device is a custom design

device which only suits a particular patient because each patient has different amounts of pitch and periodicity information (which determines the F_0 of a speech).

In this study the pitch feature is extracted using a scaled harmonic filter algorithm. The flow diagram of the algorithm is shown in Fig. 2. According to the diagram, firstly speech in .wav is used to obtain the initial values of fundamental frequencies, or referred as F_o^{raw} ; it can be obtained by pitch-tracking manually or by using available speech-related applications. Then this F_o^{raw} is fed into the pitch optimization algorithm and yield to an optimized pitch, F_o^{opt} [12]. This F_o^{opt} is used as an added input feature to the DTW-FF feature described in Step 2.

Pitch optimization is performed to resolve glitches in voice activity and pitch discontinuities due to octave errors. The algorithm of the pitch optimization is described in detail in [12]. The algorithm find the optimum pitch value for a particular time by minimizing the difference between the calculated and the measured smearing of the spectrum due to the window length.

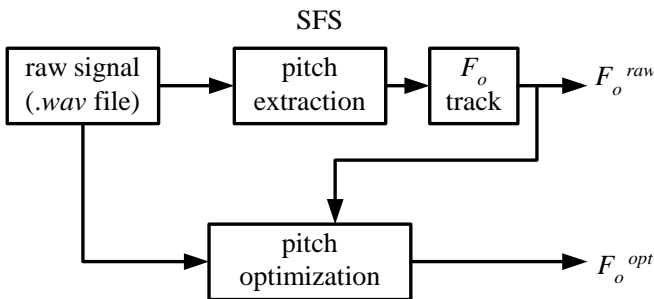


Fig. 2 Pitch feature extraction flow diagram

4. Experiment

The experiments consist of 2 phases. They are experiments of BPNN speech recognition comparing between using DTW-FF feature only and using DTW-FF feature combined with the pitch feature. Earlier experiment has showed that the DTW-FF feature does not alter the information contains in the speech since it showed no difference in its recognition rate when compared with the usage of LPC coefficients using the traditional DTW recognition engine. The result of Phase 1 and Phase 2 experiments are presented and discussed in the next section.

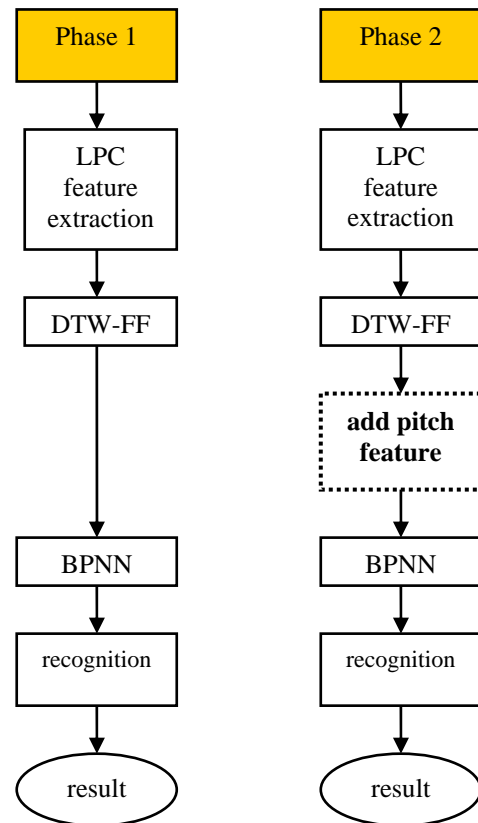


Fig. 3 Phase 1 and Phase 2 of the experiments

5. Results and Discussion

Phase 1: BPNN with DTW-FF Feature

TABLE 1: Comparison of using DTW-FF coefficients in DTW vs. BPNN

Subjects	1	2	3	4	5	6
DTW (%)	92	92	90	94	90	96
BPNN (%)	94	100	96	100	96	100

This experiment is to show that there is an improvement when using BPNN compared to the traditional DTW recognition engine. On average, the improvement from using DTW-FF in typical DTW to using DTW-FF in back-propagation neural networks is increased by 5.34 % for this particular set of experiment using 50 utterances by each subject.

DTW-FF features are obtained from the matching process in the DTW-FF algorithm. The scores have been scaled down from LPC coefficient which is a 10-order feature vector, into a coefficient (which is called as DTW-FF coefficient) derived from each frame. Besides fixing to equal number of frames between the unknown input and the reference, this activity have also tremendously reduced the amount of inputs presented into the back-propagation neural networks. Calculation to the input size reduction for example for 50 samples of 49 frames with LPC order-10 is as follows:

For input using the LPC coefficients,

$$\begin{aligned} \text{Input}_{LPC} &= \# \text{ utterance} \times \# \text{ frame/utterance} \times \# \text{ coeff/frame} \\ &= 50 \text{ utterance} \times 49 \text{ frame/utterance} \times 10 \text{ coeff/frame} \\ &= 24500 \text{ coefficients} \end{aligned}$$

For input using the local distance score,

$$\begin{aligned} \text{Input}_{LD} &= \# \text{ utterance} \times \# \text{ frames/utterance} \times \# \text{ coeff/frame} \\ &= 50 \text{ utterances} \times 49 \text{ frames/utterance} \times 1 \\ &\text{coeff/frame} \\ &= 2450 \text{ coefficients} \end{aligned}$$

Therefore, the percentage of number coefficients reduced is

$$\# \text{ coefficient reduced} = \frac{\text{Input}_{LPC} - \text{Input}_{LD}}{\text{Input}_{LPC}} \times 100\% = 90\%$$

However, this percentage is slightly higher if higher LPC order was used. For example, if LPC of order 12 is used, the percent reduction is 91.7. These means a lot of network complexities and amount of connection weights computations during the forward pass and backward pass can be reduced. Thus a faster convergence is achieved and this also allows more parallel computing of the speech patterns can be done at a time (more patterns can be fed into the neural networks at the same time).

Phase 2: BPNN with DTW-FF and Pitch Feature

TABLE 2: Recognition percentage before and after pitch feature is added to DTW-FF feature

Subjects	1	2	3	4	5	6
Before (%)	84	100	86	72	80	100
After (%)	100	100	95.9	98	90	100

The statistical test, called as T-Test has been conducted to the data in Table 6-3. This test assesses whether the means of two groups are statistically different from each other.

The formula of the T-Test, t is:

$$t = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\frac{\sigma_T}{n_T} + \frac{\sigma_C}{n_C}}} \quad (5)$$

where subscripts T and C represent the groups of data and n is the number of data in the group.

The hypothesis is set such that: $H_0: \mu_{\text{before}} = \mu_{\text{after}}$ and $H_1: \mu_{\text{before}} < \mu_{\text{after}}$. From the test, it was found that the value of t for DTW-FF in traditional DTW and in BPNN is -2.571 and -3.1247 respectively with a level of significance of $\alpha = 0.05$, this implied that $\mu_{\text{before}} < \mu_{\text{after}}$. Thus the results

rejected the null hypothesis of $\mu_{\text{before}} = \mu_{\text{after}}$. Since H_1 is true where $\mu_{\text{before}} < \mu_{\text{after}}$, then it can be concluded that the percent improvement of using DTW-FF coefficients from typical DTW to using BPNN is significant. This also implies that the back-propagation neural network is a better choice for speech recognition for this particular set of data.

On the other hand, bear in mind that a lot of network complexity and amount of connection weights computations during forward and back pass have been reduced, thus faster convergence is achieved.

6. Summary

The frame alignment based on DTW method for pre-processing of linear predictive coefficients into a new form of compressed data called DTW-FF coefficients as input into back-propagation neural networks are described in this paper. The back-propagation neural network is used as the back-end speech pattern recognition engine. Having DTW-FF algorithm, frame matching is performed to fix the different frame numbers into a suitable desired frame number. The output of the frame fixing process, which is the local distance scores are then retained because these scores later are used for recognition using the back-propagation neural networks.

From the experiments, it has been proven that DTW can be modified to suit the needs of a particular situation or application as a front-end processing of speech recognition for back-propagation neural networks, although DTW itself is a back-end recognition engine. This is an alternative method found to resolve the problem of data feeding into neural network algorithm or other subsequent pattern matching using the well known dynamic programming method. The introduction of DTW-FF coefficients in to the back-propagation neural networks also would be a good sign of fast parallel processing. This is important especially when there are a lot of data sets need to be processed at the same time so that the recognition can be obtained simultaneously.

The DTW-FF coefficients were compared to the LPC coefficients using typical DTW algorithm to identify whether or not any loss of information has occurred. From the experiments, it has been proved that there were no changes in the recognition rate, so we conclude that there is no loss of information during the frame fixing.

Pitch contains spectral information of a particular speech and this is the feature that is being used to determine the fundamental frequency, F_0 . The result of the experiment showed improvement on the recognition rate compared to only using the DTW-FF coefficients, this gives good sign how important pitch is when combined with other feature like the DTW-FF feature beside its least significant when it being used alone in speech recognition.

7. Conclusion

Initial observation from the experiment conducted leads to a resolution that the DTW-FF algorithm is able to produce a better way of representing input features into the

neural networks. These have been proven that the reformulation of the LPC feature into DTW-FF feature coefficients do not affect the recognition performance even though the coefficients size is reduced by 90% for an order 10 of LPC. As a consequence, the computation cost and network complexity have been greatly reduced, but still gain a high recognition rate than the traditional DTW itself. Therefore, this is a new approach of feature representation and combination that can be used into the back-propagation neural networks.

A higher recognition rate is achieved when pitch feature is added to the DTW-FF feature. It can be concluded that even though pitch itself cannot provide a good recognition, eventually it can be an added feature to another very reliable feature to form a very good recognition.

References

- [1] Sakoe, H. and S. Chiba. 1978. Dynamic Programming Algorithm Optimization for Spoken Word Recognition, *IEEE Transactions on Acoustics, Speech and Signal Processing*. ASSP-26(1): 43-49.
- [2] Chen, W.Y., Chen, S.H., and Lin, C.J. 1996. A Speech Recognition Method Based on the Sequential Multi-Layer Perceptrons. *Neural Networks*, Vol. 9(4): 655-669.
- [3] Abdulla, W. H., D. Chow and G. Sin. 2003. Cross-Words Reference Template for DTW-based Speech Recognition System. *IEEE Technology Conference (TENCON)*. Bangalore, India, 1: 1-4.
- [4] Rabiner, L. and B. H. Juang. 1993. *Fundamentals of Speech Recognition*. Englewood Cliffs, New Jersey: Prentice Hall.
- [5] Creany, M. J.. 1996. *Isolated Word Recognition using Reduced Connectivity Neural Networks with Non-Linear Time Alignment Methods*. University of New Castle-Upon-Tyne: Ph.D. Thesis.
- [6] Kuhn, M. H., H. Tomaschewski, and H. Ney. 1981. Fast nonlinear Time Alignment for Isolated Word Recognition. *Proceedings of ICASSP*. 6: 736-740.
- [7] Ahmadi, M., N.J. Bailey, and B.S. Hoyle. 1996. Phoneme Recognition using Speech Image (Spectrogram). *3rd International Conference on Signal Processing*. Vol 1: 675-677.
- [8] Abdul Aziz, M. A.. 2004. *Speaker Recognition System Based on Cross Match Technique*. Universiti Teknologi Malaysia: Master Thesis.
- [9] Tsai, H. L. and Lee, S. J. (1997 October). A Neural Network Model for Spoken Word Recognition. *IEEE International Conference on Systems, Man, and Cybernetics*. 5: 4029-4034.
- [10] Sudirman, R., S. H. Salleh, and S. Salleh (2006). Local DTW Coefficients and Pitch Feature for Back-Propagation NN Digits Recognition, *Proceeding of IASTED International Conference on Networks and Communications*, Thailand. 201-206.
- [11] Sudirman, R., S. H. Salleh, and T. C. Ming. 2005. Pre-Processing of Input Features using LPC and Warping Process. *Proceeding of 1st International Conference on Computers, Communications, and Signal Processing*, Kuala Lumpur. 300-303.
- [12] Jackson, P. J. B and C. H. Shadle. 2001. Pitch-Scaled Estimation of Simultaneous Voiced and Turbulence Noise Components in Speech. *IEEE Transactions on Speech and Audio Processing*. 9(7): 713-726.

Proposed Mechanism of Indigenously Electromechanical Dual Acting Pulleys CVT (EMDAP CVT) for Automotive Application

S. Ariyono*, K B Tawi, B Supriyo**, N Abu Husain

*Drivetrain Research Group
Automotive Development Center
Faculty of Mechanical Engineering
Universiti Teknologi Malaysia (UTM)*

Abstract

Almost all of the present automotive CVTs utilizing Van Doorne metal pushing V belt (VDMPVB) concept uses hydraulic actuator with single acting pulley mechanism. The hydraulic system needs continuous power to maintain the desired ratio, hence introducing power losses. The single acting pulley system operates in such a way that only one pulley sheave in each shaft moves axially and technically creates belt misalignment. This paper proposes new electromechanical system with dual acting pulley system. The electromechanical system is designed to overcome the hydraulic power loss and hence improve CVT efficiency. The basic idea of dual acting pulley movement is to eliminate belt misalignment due to ratio change. The system enables the both pulley sheaves move axially in each shaft, hence keeping the metal belt align in the center of shafts. This paper also presents several linkage mechanisms driven by DC motor to move pulley sheaves axially along the shaft. The effectiveness of each linkage mechanism will be discussed to select the suitable mechanism system for the EMDAP-CVT. Up to now, the Drivetrain Research Group, UTM has successfully manufactured three EMDAP-CVT prototypes, each has a different linkage mechanism. The experimental results show that the power screw mechanism has better performance than the others. The components of power screw mechanism can be designed symmetrically, hence reducing manufacturing cost and keeping the maintenance easier.

Key word: electromechanically system, dual acting pulleys, CVT, indigenous transmission

1. Introduction

Feasibility studies on CVT for automotive applications have long been performed because this transmission had the potential to be an ideal intersection between engine and drivetrain. Because of this potential, research and development activities on CVT for automotive applications are expanding rapidly.

In Universiti Teknologi Malaysia (UTM), Drivetrain Research Group (DRG), sponsored by the Ministry of Science and Technology (MOSTE, Malaysia, is also embarking into this field of research. Recently, this group has been involved in designing and developing an automatic electromechanical controlled dual acting metal pushing V-belt CVT gearbox for automotive applications named EMDAP-CVT. This system belt drive CVT is hoped to be an alternative to the current automatic CVT gearboxes in the automotive market.

Currently, almost all metal pushing V-belt (MPVB) drive CVT cars are equipped with semi-automatic and automatic gearboxes that are hydraulic controlled. These in turns usually resulted in more expensive cars and drivers have less control on driving the vehicles. Brace (1999) in his studies stressed out that “the fun of driving” is less with cars using automatic gearboxes. Hence, DRG tried to take this challenge and developed an electromechanically controlled MPVB CVT. Furthermore, electromechanically controlled gearboxes are more robust, lesser number of components and can be coupled with the present dry friction clutch systems which are readily available in the automotive market.

Also, most of the current MPVB CVTs use single acting pulley mechanism. This resulted in belt misalignment which, according to Tawi (1997), will affect the transmission efficiency. Thus, DRG tried to introduce dual acting pulley mechanism to overcome this phenomenon. Results demonstrate that this system has a very good potential for automotive applications. The dual acting mechanism developed by the team is now at its documentation stage for patent application.

* Permanent address: Teknik Mesin, Politeknik Negeri Semarang
Jln. Prof. H. Sudarto, S.H. Tembalang, Semarang 50329, Indonesia
Email: abusaifani@yahoo.co.uk

** Permanent address: Teknik Elektronika, Politeknik Negeri Semarang
Jln. Prof. H. Sudarto, S.H. Tembalang, Semarang 50329, Indonesia
Email: bsupriyo7763@yahoo.com

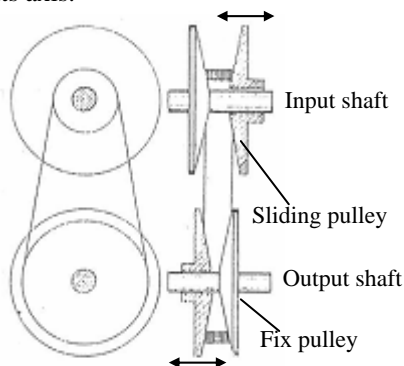
2. Hydraulic losses

Most belt type CVTs on the market are actuated hydraulically. In these CVTs, oil is pumped into cylinders on the pulley shafts to move the pulley sheaves and clamp the belt. To provide oil pressure and flow, a hydraulic pump is most often coupled to the engine crankshaft, either directly or with a fixed transmission ratio. Cylinder actuation pressures are usually controlled electronically by means of solenoid valves.

Hydraulic power loss is one of the main losses in the belt type CVT (T. Ide, 1999). The pump that provides oil pressure and flow is dimensioned for situation when maximum shifting speed and therefore high oil flow is demanded while the engine is idling and pump speed is low. Higher engine speeds thus lead to excessive oil flow, which are drained to sump. The oil still has to be kept pressurized to provides belt-clamping force. This drainage of pressurized oil is direct power loss that negatively influences the vehicle's fuel efficiency. Particularly at partial load, this loss is significant.

2.1 Misalignment

The movements of the pulleys determine the speed ratio of the CVT gearbox. Almost all CVT used in the automotive industries has one sheave of pulley moves along its axis.



It is designed in such a way that the movable pulley moves along its axis in the opposite direction. Figure 1 illustrates the movement of pulley sheave.

One input pulley sheave, considered on right side, move along on its axis, and one output pulley sheave, considered on left side moves on its axis hence creates difference pulleys ratio and unavoidably will introduce belt misalignment, Figure 2. This system is called single acting system.

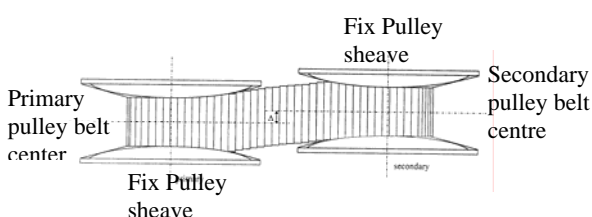


Figure 2 Belt Misalignment Model [K.B. Tawi, 1997].

3. Electromechanical mechanism

Micklem. J. D. (1994) & Tawi K. B. (1997) reveal that the compression force required to change the MPVB CVT pulley ratios are quite enormous approximately about 15 to 20 kN. Credits should be given to the designers who design the CVT pulleys mechanisms; where these tremendous amounts of forces cancel out each other and are fully taken by the pulleys' shaft. This probably explains why most of present CVT pulley systems opt to hydraulic and end up automatic.

What DRG is trying to do is, to come out with electromechanically controlled MPVB CVT as an alternative to the present automatic CVT gearboxes. The biggest challenge the team found is to come out with a mechanical mechanism that is capable of compressing the approximately 15 to 20 kN pulley axial force to change the CVT pulley ratios.

These design requirements finally, lead the team to propose power screw mechanism couple with appropriate geartrain. The idea of power screw is based on mechanical jack, which uses square thread screw as power source. Square threaded screw was chosen to distribute rotation torque into axial movement of input / output pulleys resulting ratio change. It is considered easy to manufacture rather than using surface cam. With this electromechanical linkage, it is possible to keep belt on its center line hence eliminate belt misalignment.

3.1 Dual Acting Pulley Mechanism

In order to achieve the appropriate pulley axial displacement for the CVT gearbox, there were two methods that have been considered during the design stage. That is, single and dual acting systems. For a dual acting pulley system, it has to be designed in such away that it can eliminate belt misalignment. This can be achieved by allowing both the primary and secondary pulleys to axially move inwards and outwards, respectively and accordingly. Thus in turn, provide the required CVT gearbox speed ratios, as shown in Figure 3.

Dual acting system design reduces the MPVB misalignment to zero. Hence, should lengthen belt life, improve torque transmission and should increase transmission efficiency accordingly. The design concept proposed by the team enables the CVT gearbox to be manually and variably controlled; and the cam relatively small rotational movement (to axially and continuously vary the pulley ratios) can be achieved through standard linkage mechanism, worm and worm gear set powered by simple electric motor, can be attached to the cam mechanisms; thus control the CVT gearbox speed ratios.

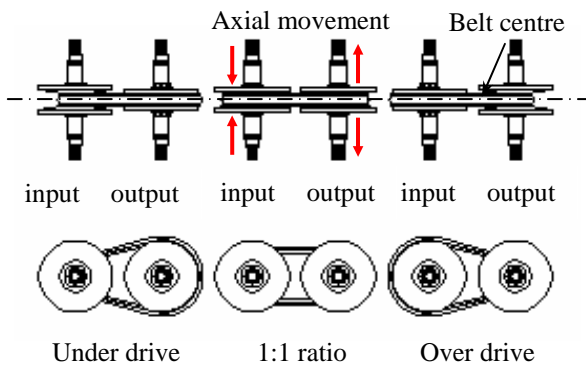


Figure 3 Basic concept of dual acting system CVT, Drivetrain Research Group ADC, UTM

3.2 Shaft and Pulleys

Figure 4 shows the pulleys and shafts designed by the team. The ball and groove splines concept allows the shaft and pulleys to be rotated together and at the same time slides axially in an out along the shafts.

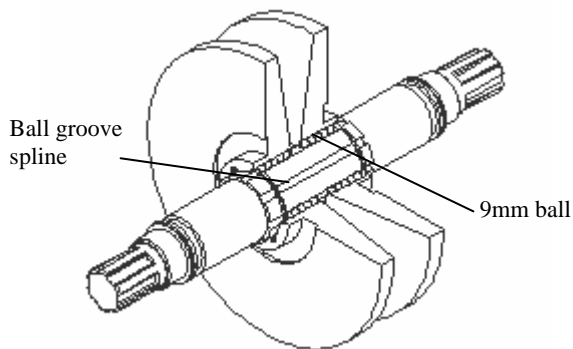


Fig. 4: Design of pulleys and shaft

Both primary (driving) and secondary (driven) shaft and pulleys set are similar in design. Their dimensions same except for the pulleys diameters. This tremendously reduces the cost of manufacturing these components. With appropriate sizing of the driving and driven pulley diameters the required 3.16 - 0.71 pulley ratios suitable for 1.3L engine can be achieved. Usually, the ratio that can be achieved using same size of pulleys with same centre distance are from 2.5 to 0.5. This is not suitable for 1.3L engine application. In addition, the difference between the driving (smaller) and driven pulleys diameter, make it possible for the MPVB to be replaced quite easily without disassemble the whole gearbox as in most of the current automotive automatic CVT gearboxes.

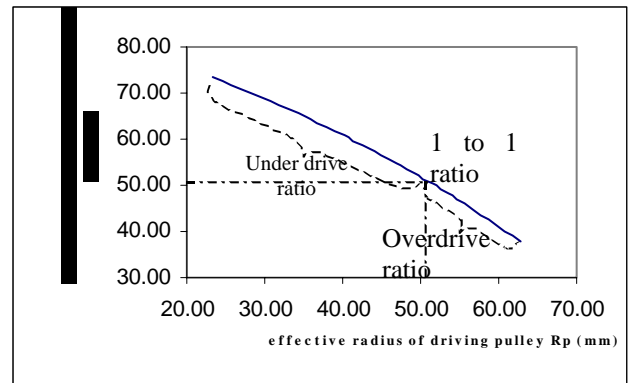


Figure 5. : Theoretical analyses result. A relation among belt effective radius at driving and driven pulleys, cam rotational and speed ratio

Figure 5 shows the characteristics of DRG CVT theoretically. The figure predicts the effective radius between driving pulley and driven pulley. It shows that the line is not straight line, i.e. the change of belt radii in the driving pulley is not perpendicular to the change of belt radii in the driven pulley.

3.3 Power Screw

In DRG automatic CVT, power screw mechanism is used for axial loading mechanism beside to move pulley axially. The movements of the pulleys determine the speed ratio of the CVT. During the speed ratio changes, the movement of the driving and the driven pulley are different forms each other to maintain the belt tension.

Both driving (primary) and driven (secondary) moving screws have a mean diameter d_m and will produce a small linear displacement, δl for a small angle of rotation, $\delta\theta$, when loaded by axial compressive force, F . Figure 6.

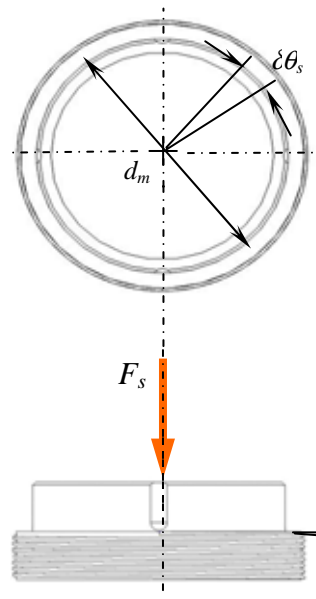


Figure 6. Moving screw

Figure 7 represents the summation of all the forces acting upon the moving screw thread. The friction force is the product of the coefficient of friction μ with the normal force, N and acts to oppose the motion.

The system is in equilibrium under the action of these forces, and hence, for primary moving screw lowering the load;

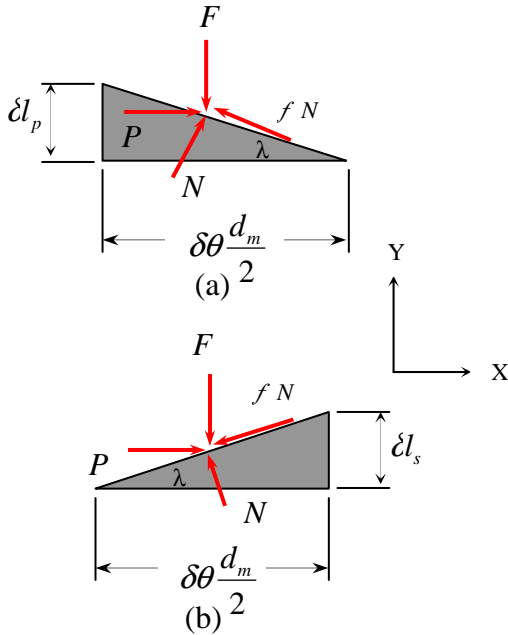


Figure 7 Force diagram; (a) lowering the load (b) lifting the load

$$\sum F_H = P + N \sin \lambda_p - \mu N \cos \lambda_p = 0 \quad 1$$

$$\sum F_V = -F + N \cos \lambda_p + \mu N \sin \lambda_p = 0 \quad 2$$

In similar manner, for secondary moving screw raising the load;

$$\sum F_H = P - N \sin \lambda_s - \mu N \cos \lambda_s = 0 \dots 3$$

$$\sum F_V = -F - N \sin \lambda_s + \mu N \cos \lambda_s = 0 \quad 4$$

Since the normal force, N is not interested, eliminate it from each of these sets of equations and solve the result for P . For raising the load this gives

$$P = \frac{F(\sin \lambda_{p,s} + \mu \cos \lambda_{p,s})}{\cos \lambda_{p,s} - \mu \sin \lambda_{p,s}} \quad 5$$

and for lowering the load,

$$P = \frac{F(-\sin \lambda_{p,s} + \mu \cos \lambda_{p,s})}{\cos \lambda_{p,s} + \mu \sin \lambda_{p,s}} \quad 6$$

From the Figure 7, where $\tan \lambda_{p,s} = \frac{2\delta l_{p,s}}{\delta\theta.d_m}$ and equation 5

& 6 are divided with $\cos \lambda_p$, then the torque required to turn the power screw for raising the load the equation can be written as:

$$T = \frac{F.d_m}{2} \left(\frac{2\delta l_{p,s} + \mu\delta\theta.d_m}{\delta\theta.d_m - 2\mu\delta l_{p,s}} \right) \quad 7$$

The torque required to lower the load is found from

$$T = \frac{F.d_m}{2} \left(\frac{\mu\delta\theta.d_m - 2\delta l_{p,s}}{\delta\theta.d_m + 2\mu\delta l_{p,s}} \right) \quad 8$$

The purpose of this analysis is to observe the effect of changing CVT speed ratios from underdrive to overdrive position and vice versa upon the required torque. This calculated torque should be supplied (by DC motor) to fixgear thus by power screw action, it will prompt the moving screw to move outwards and inwards respectively.

In this automatic EMDAP-CVT, there are two pairs of moving screw, which are purposely fabricated with different thread turn direction (right/left hand turn) in order to obtain an opposed movement between primary and secondary pulleys hence our desired speed ratio range can be achieved.

3.4 Synchronies rotation DC motor

The electromechanical dual acting pulley CVT (EMDAP CVT) system utilizes two DC servomotors ac actuators. Output shaft of the DC servomotor is directly connected to a series of gear reducers and power screw mechanism to initiate the axial pulley sheaves' movement. Based on this arrangement, the DC servomotor enables to control the sheaves' movements directly. A Van Doorne's metal pushing V-belt is placed between pulley sheaves, and runs on the surface of the sheaves. This metal belt connects the primary and secondary to transmit the power and torque from the input to the output shaft by means of friction between the belt and the pulley contact (Kanehara, S et al, 1996, Fushimi, Y et al, 1996). The contact radius of the belt with primary pulley (R_p) and the contact radius of the belt with secondary pulley (R_s) determines the transmission ratio. The closer the distance between the movable sheaves, the bigger the contact radius will be. The primary and secondary parts of the CVT system have the same components. The components of each part (primary or secondary) consist of a dc servo motor, a gear reducer with ratio of (30:1), a gear reducer with gear ratio of (60:14), power screw mechanism, and two movable metal pulley sheaves for clamping the metal pushing belt. The block diagram of the EMDAP-CVT system can be seen in figure 8. In laboratory experiments, parameters that can be directly measured are X_p (axial position of primary pulley), X_s (axial position of secondary pulley), ω_p (angular speed of primary pulley), ω_s (angular speed of secondary

pulley), T_p (torque of primary pulley) and T_s (torque of secondary pulley).

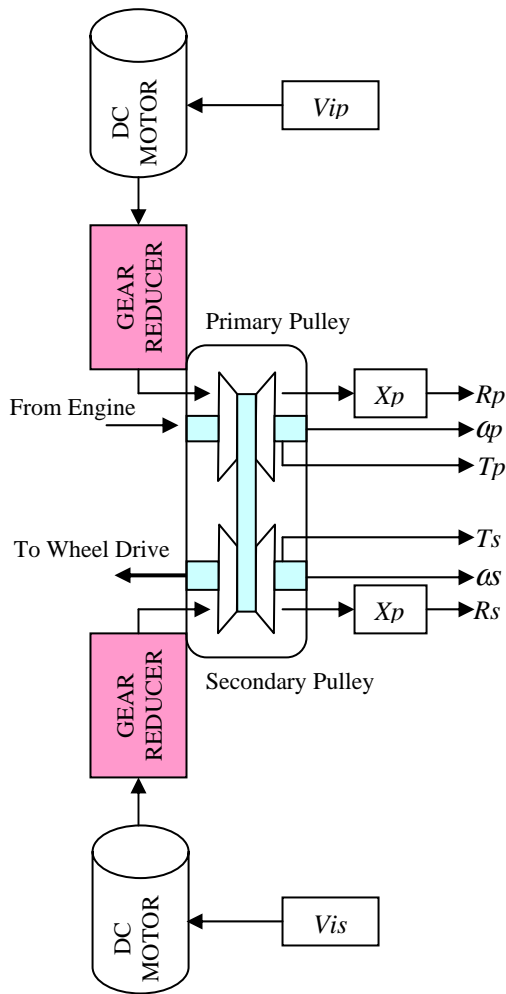


Figure 8. The block diagram of EMDAP-CVT

The primary motor actuates the primary pulley movement for transmission ratio change, while the secondary motor actuates the secondary pulley movement for clamping force. A spring disc is inserted in the back of each secondary pulley sheave to provide continuous clamping force to the belt, and to reduce excessive slip during transmission ratio change. When the CVT is on an underdrive position, the primary belt radius is minimum while the secondary belt radius is maximum. When the ratio change is called, the primary motor will actuate the primary pulley axially to the new value of primary radius, and at the same time the secondary motor will actuate the secondary pulley axially to provide the optimal clamping force for preventing belt slip. These movements will stop if the desired ratio is achieved. When the CVT is on the overdrive position, the primary belt radius is maximum, while the secondary belt radius is minimum.

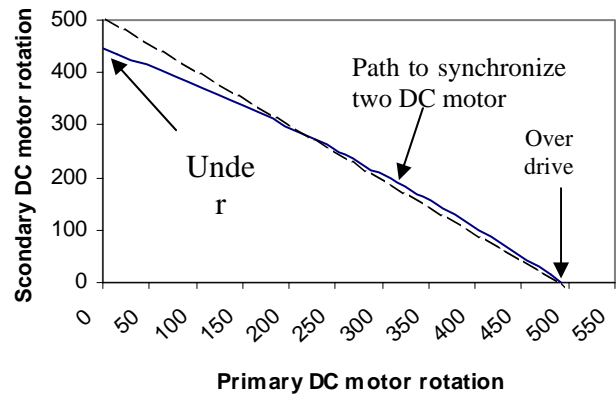


Figure 9. Synchronizing DC motor to prevent belt

Figure 9 shows that rotation of primary DC servomotor is not equivalent to the rotation of secondary DC servomotor, i.e. dash line. As seen in figure 5 that the effective radii of the each pulley are not linear, the distance achieved by the primary and secondary pulley to clamp the metal belt is not the same. When the primary pulley move toward the metal belt to create new ratio x mm, the secondary pulley will not move x mm outward metal belt.

4. Technical description

In general, this CVT gearbox (Figure 10) consists of an input and an output shafts. The input shaft holds the driving pulleys and a pair of power screw mechanisms with the shaft as their axis of rotation. Similarly, the output shaft holds the driven pulleys and a pair of power screw mechanisms with the shaft also, as their axis of rotation. Each of the pairs either the primary or secondary power screw mechanisms was connected accordingly by the gear linkage system to the DC servomotor. To ensure that the metal pushing V-belt (MPVB) remain taut during operation, two DC servomotor has to follow path in figure 9.

Torsional loading from the engine enters the CVT gearbox via the input shaft. The input shaft which is balls splined to both the driving pulley sheaves, convert this torsional load into pushing forces to the MPVB via the effective contacts of the circumferential cone surfaces of the driving pulleys and the sides of the MPVB segments sides, as circumference tangential forces. These circumference tangential forces then, compress the MPVB segments and at the same time push them forward. These pushing forces are then transferred to the output shaft in a similar manner but the force flow path is in the reversed direction as that of the driving pulleys. That is, the pushing forces of the MPVB are being converted into torsional load again to the driven pulley sheaves cones via the circumferential effective contacts of the MPVB segments sides and the pulleys halves cones surfaces, as

circumferential forces. This torsional load from the driven pulleys sheaves then torsioned the output shaft via the sliding ball splined contacts.

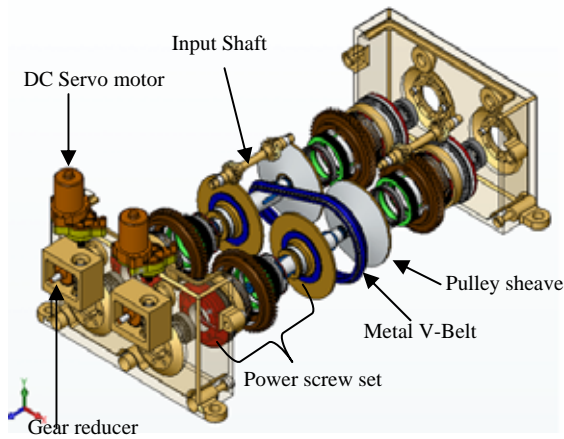


Figure 10. Exploded view of EMDAP-CVT

The two main tasks of the power screw mechanisms then, are to ensure that the belt remain taut during operation and are at the required pulley ratio. The appropriate pulley ratios are being achieved automatically by the synchronizes intelligent control system. The gear ratio is obtained from the macroscopic controller, which control the relation between the engine and drive train by using intelligent controller. The intelligent controller will select the proper gear ratio to keep the engine speed constant at its trajectory (S. Ariyono at al, 2006). The ratio obtained then is fed to the microscopic controller by the synchronized intelligent controller (B. Supriyo, 2006).

The experimental of this indigenously electromechanical dual acting CVT shows work well. It can transmit optimum torque from the 660 cc ICE engine and run at almost maximum engine speed.

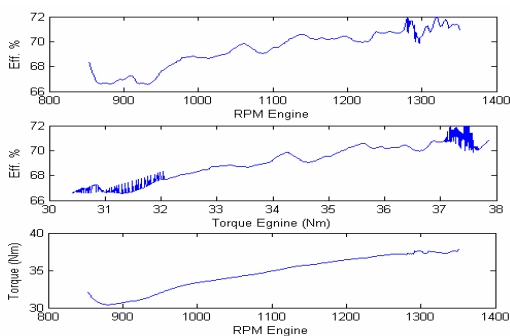


Figure 11. Performance of EMDAP-CVT at overdrive ratio (0.6 : 1)

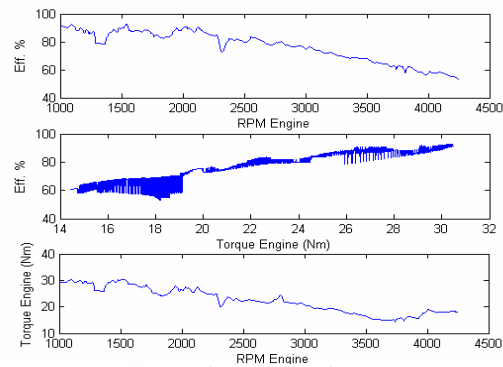


Figure 12. Performance of EMDAP-CVT at underdrive ratio (2.2 : 1)

5. Conclusion

This paper intends to explore the possibility to design an indigenously electromechanical dual acting CVT gearbox for automotive applications. This paper has presented the results of the feasibility study, discussed the basis of the concept selected and the detail mechanical layout of the CVT that has been developed by DRG. The proposed mechanism shows work well when the EMDAP CVT was tested its performance with the 660 cc ICE engine. The efficiency can lies 65% - 80% and the prototype can deliver maximum torque of the engine. Figure 11,12 shows the under drive and overdrive performance of the prototype.

6. References

- B. Supriyo, KB Tawi, H. Jamaluddin S. Ariyono, (2006), "Dc Motor Position Control For Pulley Axial Movement Of An Electromechanical Dual Acting Pulley (Emdap) CVT System", RIVET'06, Kuala Lumpur, Malaysia.
- Brace C. (1999), " An Operating Point Optimiser for the Design and Calibration of an Integrated Diesel/CVT Power train ", Proceeding of The Institution of Mechanical Engineers Journal of Automobile Engineering (part D).
- Fushimi, Y. , T. Fujii and S. Kanehara. 1996. A Numerical Approach to Analyse the Power Transmitting Mechanisms of a Metal Pushing V-Belt Type CVT, SAE Technical Paper Series 960720.
- Kanehara, S. , T. Fujii and S. Oono.1996. A Study of a Metal Pushing V-Belt Type CVT : Macroscopic Consideration for Coefficient of Friction between Belt and Pulley, SAE Paper No. 9636277
- Micklem J. D. (1990). "Research Project On The Modelling Of The Ford CTX / Van Doorne Continuously Variable Transmission System", University of Bath.
- S. Ariyono, KB Tawi, H. Jamaluddin, B. Supriyo, (2006), "Engine Speed Control Using Artificial Neural Network for Vehicle Equipped with

- Electromechanical CVT”, RIVET’06, Kuala Lumpur, Malaysia.
- T. Ide (1999), “Effect of Power Losses of Metal V-Belt CVT Components on the Fuel Economy”, proceeding of CVT ’99 Congress, pp. 93-98, Eindhoven, The Netherlands.
- Tawi, K.B. (1997). Investigation of belt misalignment effects on metal pushing v-belt continuously variable transmission. Ph.D thesis, Cranfield University.

Thermal Behaviour of Coal /Biomass Blends During Co-Pyrolysis

Ahmad Hussain, Farid Nasir Ani, Amer Nordin Darus

Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310, Skudai, Johor

Tel: +607-5534785, Fax: +607-5566159, Email: ahmadutm@gmail.com

ABSTRACT

Co-firing of coal and biomass fuels is being presently considered as an alternative means of reducing CO₂ emissions. Besides as single fuel, it is also considered sensible to utilise biomass in co-combustion in existing firing systems, such as coal-firing power stations. Whereas the construction of new decentralised plants fired exclusively with biomass would require high investment costs, the retrofit for co-combustion in an existing coal-firing plant requires less supplementary investments. However, the application of co-combustion technologies involving biomass and coal for power generation requires a proper understanding of the thermal properties and reaction kinetics of coal/biomass blends. The information pertinent to thermal events and kinetics play an important role in the efficient design, operation, and modelling of co-fired boilers. The thrust of this paper therefore was to identify the thermal events taking place during pyrolysis of coal/biomass blends and obtain kinetic data based on thermogravimetric analysis (TGA).

Investigations into the thermal behaviour during co-pyrolysis of coal, biomass materials and coal/biomass blends prepared at different ratios (10:90, 20:80, 30:70 and 40:60) have been conducted using a TGA apparatus. Coal sample selected was Mukah sub-bituminous coal from Sarawak while palm shell waste (PSW) was used as biomass material. Three thermal events were identified during the pyrolysis. The first two were dominated by the biomass pyrolysis, while the third was linked to the coal pyrolysis, which occurred at much higher temperatures. No interactions were seen between the coal and biomass during co-pyrolysis. The pyrolytic characteristics of the blends followed those of the parent fuels in an additive manner. Among the tested blends, 40:60 blends showed the lowest activation energies of 80.8 KJ/mole for coal/WW blends. Besides the pyrolysis of coal alone, the 10:50 coal/biomass blends had the highest reaction rate of $5 \times 10^3 \text{ sec}^{-1}$.

Keywords: Activation energy, biomass, co-firing, co-combustion, reaction rate, thermogravimetric analysis

1. Introduction

Malaysia has a coal mining history dating as far back as 1851 when the first coal mine was opened at Labuan. In fact the known coal resources in the country were swiftly dealt with by describing it as of "low quality and of little potential". The present coal resource figure stands at 1,025 million tonnes; 725 million tonnes are found in Sarawak, 282 million tonnes in

Sabah and 17 million tonnes in Peninsular Malaysia. The coal fields are identified as Merit Pila, Bintulu and Mukah Balingian in Sarawak and Malibau coal areas in Sabah. The coal areas of the country are shown in Figure 1. Coal quality ranges from the lignite rank to anthracite; sub-bituminous coal, however, predominates. As a result of the oil crises of 1973 and 1979, it was thought inevitable for energy diversification from oil

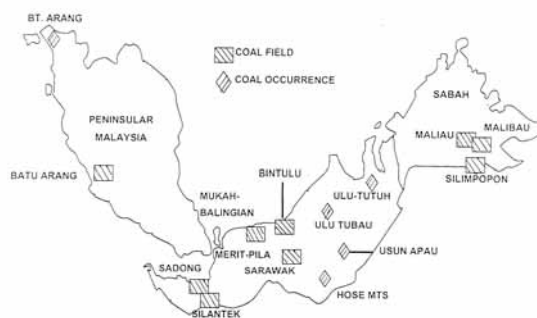


Figure 1: Locations of Coal reserves in Malaysia [1]

In 1980, oil and petroleum products contributed 87.8% of the commercial energy supply, and in 1984 the Malaysia's four-fuel strategy based on oil, gas, hydropower and coal was promulgated. This basically set the scene for greater coal use. The first 600 Mw coal-fired Sultan Salahuddin Abdul Aziz (SSAA) power station in Kapar, Peninsular Malaysia was commissioned in 1985. Since then one additional 100 MW station started operation in 1997 in Sejingkat, Sarawak.

The projected coal demand in Asia to the year 2010 have been a buoyant one, with most countries

like Korea, Hong Kong, Philippines, Malaysia, India and Thailand reporting increasing use of coal. Even major coal producing countries like Indonesia and China are projected to face shortage and their capacity for export is expected to be largely curtailed, with production barely matching local demand. Australia has been looked on to meet the increase in coal demand by stepping up production and export. The power generation mix in Malaysia is shown below.

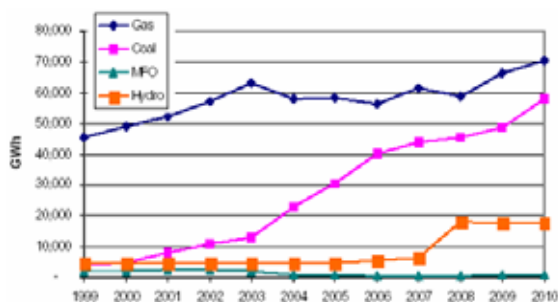


Figure 2: Power Generation Mix in GW [2]

In the background of this regional development, Malaysia was projected for steep increase in coal use in the next decade with the construction of at least 2 more cement plants and several large coal fired power stations. Within Malaysia, there is a large market for coal, and the demand is projected to rise steeply to a total of about 14 million tonnes, to the year 2010. Although it may not be possible to meet this entire demand by indigenous sources, estimate shows that it may be possible to produce 6.2 million tonnes, or 40% of the demand. The focus of this projection is that the attractive coal deposits like Mukah coal in Sarawak which have not been utilized much. This study is focused on identifying the pyrolysis behaviour of Mukah coal and its blends with oil palm shell waste.

2. Thermogravimetric analysis (TGA) of Mukah coal and its blends

In this paper, pyrolysis of Mukah coal and its blends, were carried out using thermogravimetric analysis (TGA). The effects pyrolytic temperature and heating rate on the pyrolytic properties (for example, the shape of thermograms) were investigated to determine the main factor affecting the pyrolytic process.

The samples were dried to remove free moisture absorbed. The pre-dried samples were crushed and sieved to several size fractions. The pyrolysis was carried out using a thermogravimetric analyzer (Perkin Elmer Pyris TGA). Approximately, 20 mg sample was placed in a platinum pan, which was suspended by a platinum wire. The furnace was heated from ambient temperature to a programmed temperature of 1000 °C at a constant heating rates (25 & 80 °C/min). Purified nitrogen (99.9995%

purity) at a constant flow rate was used as the purge gas to provide an inert atmosphere for pyrolysis and to remove any gaseous and condensable products evolved, thus minimizing any secondary vapor-phase interactions. The sample was heated by both radiation from the furnace wall and convection of the purge-gas rushing through the furnace chamber. The sample weight was measured continuously by a microbalance as a function of time or temperature.

3. Determination of kinetic parameters

Thermogravimetric data is used in characterizing the fuel as well as in investigating the thermodynamics and kinetics of the reaction and transitions that results from the pyrolysis of samples. TGA is useful in providing kinetic data as a function of various reaction parameters such as temperature and heating rate [3]. Currently several methods were available in the literature that can be used to calculate kinetic parameters [4]. The rate of reaction is given by:

$$\frac{d\alpha}{dt} = A e^{-E/RT} (1-\alpha)^n \quad (1)$$

where A (min^{-1}) is the frequency or pre-exponential factor of the pyrolytic process, E (J/mol) is the activation energy of the pyrolytic process, R (J/mol K) is the universal gas constant, T (K) is the absolute temperature, n is the order of reaction, t is the time, and α is the fraction of reactant decomposed at time t (min).

The fractional reaction α is defined in terms of change in the mass of the sample

$$\alpha = \frac{w_o - w}{w_o - w_f} \quad (2)$$

Where w_o , w , w_f are the initial, actual and final weights (mg) of the sample, respectively.

In order to determine the values of kinetic parameters, the integral method is used to solve Equation (1).

For constant heating rate β :

$$\beta = \frac{dT}{dt} \quad (3)$$

Equation (1) can be expressed by the following equation:

$$\frac{d\alpha}{dT} = \frac{A}{\beta} e^{-E/RT} (1-\alpha)^n \quad (4)$$

Rearranging and integrating Equation (4), the following expressions can be obtained:

$$\frac{1-(1-\alpha)^{1-n}}{1-n} = \frac{A}{\beta} \int_0^T e^{-E/RT} dT \quad (5)$$

Since $\int e^{-E/RT} dT$ has no exact integral, $e^{-E/RT}$ can be expressed as an asymptotic series and integrated, with the higher order terms ignored.

$$\frac{1-(1-\alpha)^{1-n}}{1-n} = \frac{ART^2}{\beta E} \left[1 - \frac{2RT}{E} \right] e^{-E/RT} \quad (6)$$

Expressing Equation (6) in logarithmic form

$$\ln \left[\frac{1-(1-\alpha)^{1-n}}{T^2(1-n)} \right] = \ln \left[\frac{AR}{\beta E} \left[1 - \frac{2RT}{E} \right] \right] - \frac{E}{RT} \quad (for\ n \neq 1) \quad (7)$$

If $2RT/E \ll 1$ is assumed, Equation (7) becomes

$$\ln \left[\frac{1-(1-\alpha)^{1-n}}{T^2(1-n)} \right] = \ln \left[\frac{AR}{\beta E} \right] - \frac{E}{RT} \quad (for\ n \neq 1) \quad (8)$$

If $n=1$, the following equation can be used

$$\ln \left[-\frac{\ln(1-\alpha)}{T^2} \right] = \ln \left[\frac{AR}{\beta E} \right] - \frac{E}{RT} \quad (for\ n=1) \quad (9)$$

Thus, a plot of

$$\ln \left[\frac{1-(1-\alpha)^{1-n}}{T^2(1-n)} \right] \text{ versus } \frac{1}{T} \quad (for\ n \neq 1) \quad (10)$$

or

$$\ln \left[-\frac{\ln(1-\alpha)}{T^2} \right] \text{ versus } \frac{1}{T} \quad (for\ n=1) \quad (11)$$

should result in a straight line of slope $-E/R$ for the proper value of n . The criterion used for the acceptable values of E and A is that the final value of n should yield the values of E whose linear correlation coefficient are the best.

4. Results and discussions

Figure 3 shows the residual weight fractions of Mukah coal undergoing pyrolysis for heating rates of 25 and 80 °C/min. Figure 3 also shows the thermal events observed during pyrolysis. It should be noted that each thermal event is defined as the region where the slope of TGA curve is constant. A major shift in the slope of the TGA curve was thus treated as the end of a thermal event and commencement of a new thermal event. The thermograms

shows that as the heat wave propagates into the solid then inherent moisture evaporates and this ends at about 130 °C. This is represented by 'Thermal Event 1' in Figure 3. As the temperature is further increased, the breakdown of more stable polymers begins. The thermograms showed that pyrolysis commences after about 260 °C. Subsequently, it showed a significant weight loss, indicating the occurrence of main decomposition. These are represented by 'Thermal Event 2' and 'Thermal Event 3'. Similar trends for the Mukah coal blends with oil palm shell wastes having 10% Or 20% coal are shown in Figures 4-5.

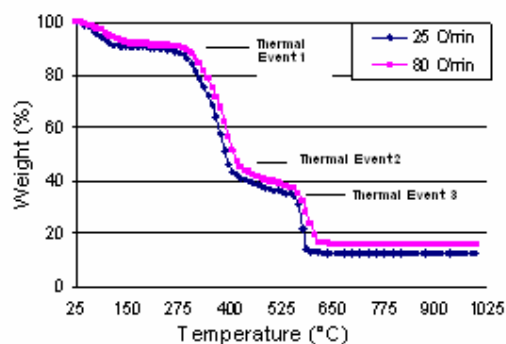


Figure 3: Variation of weight percent with time for pyrolysis of different Malaysian coals at a heating rate of 25 °C/min

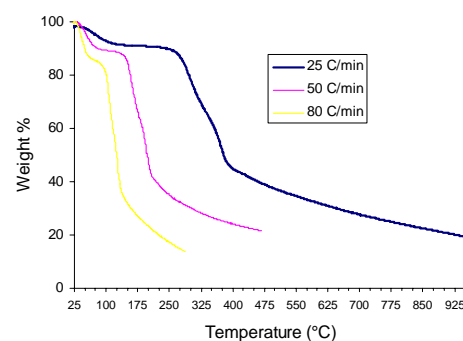


Figure 4: Variation of weight percent for pyrolysis of 10% coal blend at different heating rate

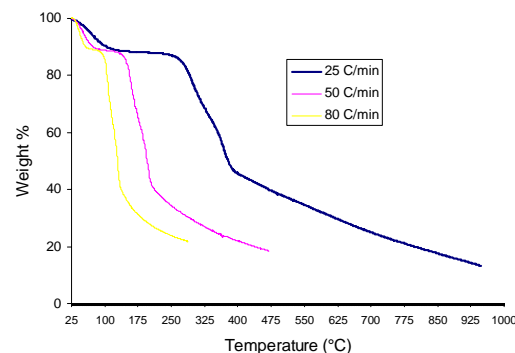


Figure 5: Variation of weight percent with temperature for pyrolysis of 20% coal blend at different heating rate

At the beginning, CO₂ and CO were released as the main gaseous products and at the main decomposition period, a large amount of gaseous products such as CO₂, CO, H₂ and hydrocarbons (i.e. CH₄; C₂H₄; C₂H₆) were released, resulting in a significant weight loss

The heating rates influence the shape of the TGA curve. When a sample decomposes then the vapour pressure of the gaseous products exceeds the ambient partial pressure. At lower heating rates the sample temperature is more uniform and diffusion of the product gases can occur through the sample but when the heating rate is increased then such free diffusion is inhibited and decomposition temperature is increased. Also it was found that at lower heating rates the decomposition atmosphere is more uniform and the decomposition reaction is completed within a narrower temperature interval.

Figure 4-5 shows the residual weight fractions for pyrolysis of coal blends at various heating rates. It could be seen that there is an obvious lateral shift in the thermograms for different heating rates. In addition, there is an effect of heating rate on the total weight loss. As the heating rate was increased, a faster pyrolytic reaction occurred, resulting in higher pyrolytic conversion to volatile species.

In Figure 6, the main decomposition for Mukah coal was essentially completed by an elapsed duration of 9 min and 15 min for heating rates of 80 and 25 °C/min respectively.

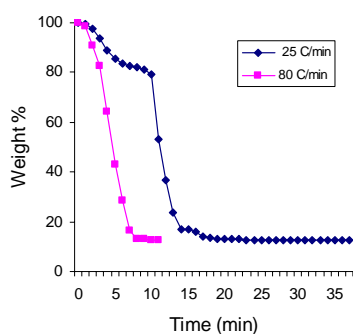


Figure 6: Variation of weight percent with time for pyrolysis of Mukah coals at different heating rates

Figures 7 show the derivative thermograms (dm/dt) for the pyrolysis of Mukah coal at various heating rates. The height of the peak at any temperature gives the rate of mass change at that temperature.

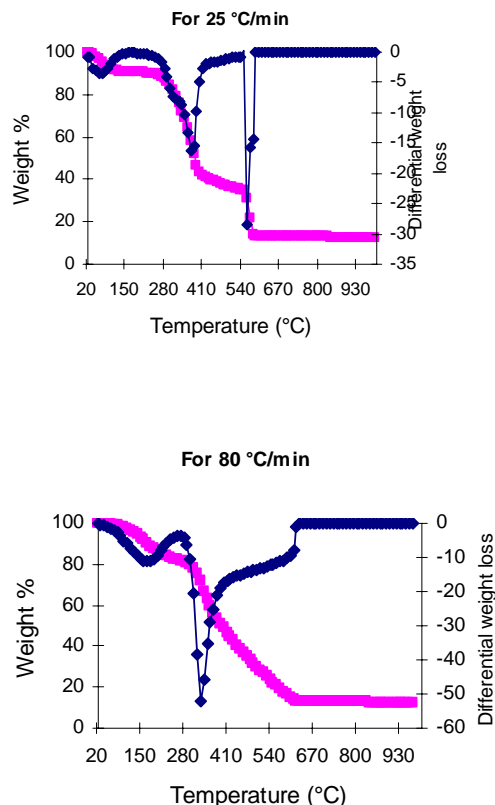


Figure 7: Differential thermogram (DTG) for pyrolysis of different Mukah coal at a heating rates of 25 and 80 °C/min

Figures 7, suggest that for the heating rates 25 and 80 °C/min the maximum decomposition rates are 30 and 50 mg/°C. The decomposition rates are identical for different particle sizes as the particle sizes below 2 mm are primarily considered as fine particles and their pyrolysis is pure reaction kinetics controlled. It could be seen clearly that there existed separate steps of reactions that took place in distinct temperature regimes with obvious maximas for different heating rates. The heating rate had influenced not only the maximum rate of pyrolysis and its temperature, but also the starting and ending temperatures for the pyrolytic process. This phenomena is also confirmed by the fact that there was a lateral shift to higher temperatures during pyrolysis as the heating rate was increased.

Using data from the pyrolysis thermograms (Figures 3-5, the kinetic parameters, the activation energy (E) and the frequency factor (A), were estimated using Figure 8 with high correlation coefficients (above 0.94) and all results are listed in Table 1.

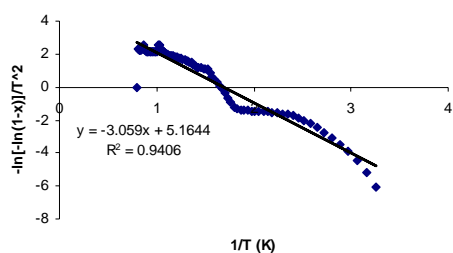


Figure 8: Determination of kinetic parameters using curve fitting

Table 1: Kinetic Parameters for the pyrolysis of Mukah coal and its blends

Fuel Type	Heating rate °C/min	Activation Energy 'E' kJ/mole	Frequency factor 'A' 1/sec
Mukah Coal	25	110.6	1.05×10^4
	80	123.6	9.4×10^3
10% Blend	25	100.8	4.9×10^3
20% Blend	25	93.3	4.2×10^3

For all the heating rates, the orders of reaction were first-order reaction mechanism for all fuels. As the heating rate was increased, the activation energy changed a bit but the frequency factor was dependent on heating rate, increasing progressively from 9.4×10^3 to $1.04 \times 10^4 \text{ sec}^{-1}$ for Mukah coal. This suggests that the higher the heating rate, the easier and faster would be the pyrolytic reaction. These parameters can be used to predict the time-conversion profiles for the pyrolytic process of different heating rates. A similar pyrolytic behaviour can be seen for 10% and 20% coal blends with oil palm shell wastes.

5. Conclusions

Based on the thermogravimetric analysis, pyrolytic temperature and heating rate were found to have significant influences on the pyrolysis of Mukah coal and its blends. The two-stage reaction characteristic was confirmed by obvious maximas in the derivative thermogram (DTG) for pyrolysis under different heating rates. DTG curves suggest that heating rates affect the decomposition rates. However, as the particle were primarily fine particles so the pyrolysis was not affected by particle size rather it was found to be pure reaction kinetics controlled. It could be seen clearly that there existed separate steps of reactions that took place in distinct temperature regimes with obvious maximas for different heating rates. Kinetic parameters (activation energy and frequency factor) were obtained by curve fitting the experimental data.

Acknowledgement

The author (AH) would like to thank Malaysian Government Commonwealth Secretariat, Public Service Department Malaysia (JPA), Putrajaya for funding the PhD program. Thanks are due to SIRIM Environment and Bioprocess Technology Centre for granting permission to use their experimental facilities.

References

- [1] Pei, C. K. (2000). The status of Malaysian coal industry. *Proceedings of the Asia Pacific Mining and Quarrying (APMQ) 2000*, Kuala Lumpur, Malaysia , 123-139.
- [2] Cox, R., Chen, S.P., Hon, V., Unya, A. and Batoi, R. (2000) . Economic geology and mineral potential of west Sarawak, Malaysia. *Proceedings of the Asia Pacific Mining and Quarrying (APMQ) 2000*, Kuala Lumpur, Malaysia , 155-174.
- [3] Ismail , A. F., Shamsuddin, A. H. and Mahdi F. M. A. (1999). Reactivity studies of rice husk combustion using TGA. *Proceedings of the World Renewable Energy Congress, Malaysia*, 44-50.
- [4] Guo, J. and Lua, A. C. (2001). Kinetic study on pyrolytic process of oil palm solid waste using two step consecutive reaction model. *Biomass & Bioenergy* 20: 223-233.

Investigation on Co-Combustion Characteristics of Palm Shell Waste and Coal in Circulating Fluidized Bed Combustor

Ahmad Hussain, Farid Nasir Ani, Amer Nordin Darus

Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310, Skudai, Johor
Tel: +607-5534785, Fax: +607-5566159, Email: ahmadutm@gmail.com

Abstract

Circulating fluidised bed (CFB) combustion is considered to be one of the latest technologies to thermally convert solid fuel into useful energy. However, in Malaysia, this is the first practical experience with this kind of combustor working on biomass and coal. This paper describes the results from a bench-scale CFB combustor (CFBC), installed at SIRIM Berhad, Shah Alam, for combustion of different fuels. The purpose of the tests was to investigate the suitability of the selected fuels for energy production using CFBC while taking care of the flue gas emissions. The measurements for temperatures and emissions were done for different samples of oil palm shell waste and its blends with sub-bituminous Sarawak coal. The concentrations of CO, NO_x and CO₂ in the flue gas were measured continuously. The effects of some parameters, such as coal/palm shell waste blending ratios, feed rate, primary air rate and excess air on the combustion characteristics and the emission of SO₂ and NO_x were clarified. The experimental results indicate that the incinerator temperature decreases with increase in primary air rate and palm shell waste/coal mass ratio. When the excess air ratio increases, the dense bed temperature increases, but the temperature in the dilute phase region first increases, and then it decreases with further increase in the excess air ratio. The NO_x emission decreases with increase in coal blending ratio. It was found that higher temperature and lower excess air ratio are beneficial to suppressing NO_x emission from incineration. It was found that the NO_x content ranged from 38-127 ppm while the CO emissions were high for some operating conditions. However, the experiments gave sufficient information on the main process and the co-combustion of coal blends was successfully demonstrated which opens a way for utilising this sustainable fuel source which is available in abundance in Malaysia.

Keywords: Biomass, co-combustion, circulating fluidized bed, emission characteristics, palm shell waste

1. Introduction

Biomass was the world's first source of fuel and energy, but when coal, petroleum and natural gas became widely available, its usage declined. However, in recent years, its interest has been renewed, much of being focused on its efficient conversion for energy. The importance of biomass energy in the developing countries today, is indisputable (Kate et al., 1997, Bhattacharya S.C., 1997). With respect to global issues of sustainable energy and reduction in greenhouse gases, biomass is getting increased attention as a potential source of renewable energy. According to the World Energy Assessment report, 80% of the world's primary energy consumption is contributed by fossil fuel, 14% by renewable (out of which biomass contributes 9.5%) and 6% by nuclear energy sources (Rogner et al., 2000; Devi et al., 2003, Bui et al., 1994).

Among the proven combustion technologies (such as grate-fired systems, suspension-fired systems, fluidised bed systems), the fluidised bed technology is reported to be the most efficient and suitable for converting agricultural and wood residues into energy (Bridgwater A.V. 1994). The emissions from biomass combustion systems, including products of complete combustion (CO₂) and incomplete combustion (CO, char particles, tar, polycyclic aromatic hydrocarbons and other organic compounds), as well as NO_x, SO₂, HCl and ash particles, are affected by the combustion method as well as by operating conditions and fuel properties (Dornburg et al., 2001). It is to be appreciated that the combustion phenomena are very complex, multidimensional, unsteady and multi-disciplinary (Pierre J., 2004).

Malaysia as being the largest producer of palm oil in the world, generates a significant amount of oil palm wastes (Ani, F.N., 1992, Jamil et al., 1999). In the year 2005 her production of palm oil is about 14.97 millions tons and the total solid wastes generated by this industry has

amounted to more than 3.96 millions tons (PORIM, 2005). In addition to the biomass wastes, Malaysia is also blessed with a large quantities of coal which is still unutilized for energy purpose. The local coals offer tremendous potential to be used for heat and power generation. Co-firing is also emerging as one the future technology for waste combustion (Rogner et al., 2000, Devi et al., 2003).

As a matter of fact, the known coal resources in Malaysia were swiftly dealt with by describing it as of "low quality and of little potential". The present coal resource figure stands at 1,025 million tonnes; 725 million tonnes are found in Sarawak, 282 million tonnes in Sabah and 17 million tonnes in Peninsular Malaysia. The coal fields are identified

as Merit Pila, Bintulu and Mukah Balingian in Sarawak and Malibau coal areas in Sabah. Coal quality ranges from the lignite rank to anthracite; sub-bituminous coal, however, predominates.

The projected coal demand in Asia to the year 2010 have been a buoyant one, with most countries like Korea, Hong Kong, Philippines, Malaysia, India and Thailand reporting increasing use of coal. Australia has been looked on to meet the increase in coal demand by stepping up production and export. The power generation mix in Malaysia is shown in Figure 1. The focus of this projection is to utilize the attractive coal deposits like Mukah coal in Sarawak which is available in abundance.

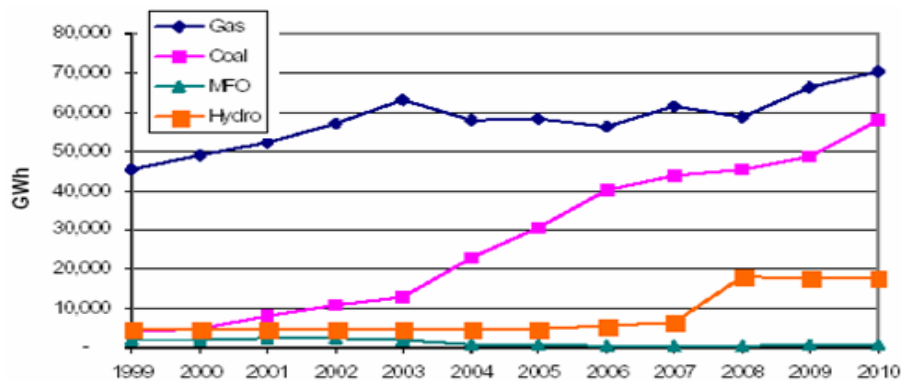


Figure 1: Power Generation Mix in GW [Cox et al., 2000]

This paper deals with the experimental study of the combustion of oil palm shell waste and its blend with Mukah coal in a circulating fluidized-bed combustor (CFBC). The main objectives of this work were to study formation and reduction of the major gaseous pollutants (CO and NO_x) in the CFBC when firing with selected blended ratio, and to see the combustion behaviour at different operating conditions.

2. Materials and Methods

In general, biomass as a fuel, is characterized by high moisture and volatile content, low bulk density, low specific energy and normally low ash content.

The fuels used for the experimental work were investigated for their proximate and ultimate analyses properties. The analyses were done at Malaysian Institute of Nuclear Technology (MINT), Bangi. The proximate analysis was done using the standard ASTM standard (D-3175) while the ultimate analysis was done using CHNS analyzer. The results of the analyses are being shown in Table 1 and Table 2.

Table 1: Proximate analysis (in wt%) palm shell waste and its coal blends on as received basis

Material	Moisture Content	Volatile Matter	Fixed Carbon	Ash
Palm Shell Waste	10.22	65.09	22.69	2.00
5% Coal Blend	10.25	63.00	24.00	2.75
10% Coal Blend	10.45	61.19	24.88	3.48
20% Coal Blend	10.50	57.50	27.25	4.25
40% Coal Blend	11.00	52.00	32.25	4.75
60% Coal Blend	12.25	41.75	36.75	9.25
80% Coal Blend	12.75	37.75	38.75	10.75
100% Coal	14.25	25.25	48.50	13.75

Table 2 : Ultimate analysis of palm shell waste and its coal blends (in wt%) on dry ash-free basis

Material	N	C	H	S	O	GHV
	%	%	%	%	%	MJ/Kg
Palm Shell Waste	0.22	45.95	5.81	-	48.02	19.59
5% Coal Blend	0.24	45.07	5.66	0.10	48.93	21.70
10% Coal Blend	0.27	46.28	5.86	0.16	47.43	22.19
20% Coal Blend	0.35	46.76	5.17	0.20	47.52	22.59
40% Coal Blend	0.34	47.70	5.56	0.32	46.08	23.38
60% Coal Blend	0.32	48.90	5.38	0.58	44.82	24.11
80% Coal Blend	0.64	52.05	5.43	0.75	41.13	24.71
100% Coal	0.45	69.45	6.37	0.98	22.75	25.68

2.1 Experimental Test Rig

The experiments were carried out at SIRIM Berhad. The experimental apparatus consists of a circulating fluidised bed using alumina particles as inert material, feeder, secondary combustion furnace, gas cooler, dust collector, blower and control panel. A schematic diagram of the test rig is being shown in Figure 2.

The CFB test rig is a vertical tubular furnace having an inner diameter of 50 mm and a height of 2000 mm. The

fluidizing air supplied from the bottom and the circulating particles are separated from air by the cyclone and the particles returned through the loop seal on the downstream side of the furnace. The rig is provided with three external electric heaters (upper, center and lower), which enable to control the temperature in the CFB. A heater is installed also on the air supply line to preheat air. For monitoring the combustion condition, the concentrations of CO, CO₂ and O₂ in the produced gas were measured using a gas analyser.

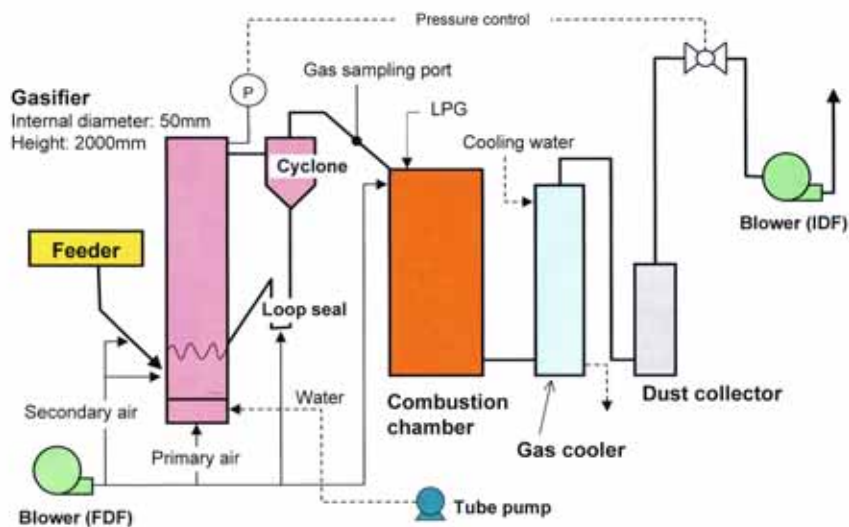


Figure 2: Layout the experimental CFB system at SIRIIM

Palm shell waste is supplied to the CFB through the screw feeder and rotary valve. The feed rate from the screw feeder can be adjusted by changing the rotational speed of the driving motor through inverter control. The rotary valve is located after the screw feeder to keep the seal of the CFB. The chute on the downstream side of the rotary valve is cooled by aeration to prevent fusing of biomass sample by overheating.

The secondary combustion furnace is designed to completely burn the combustible gases, such as CO and H₂ contained in the gas produced by CFB. The mixture of the produced gas and air is burnt by a constantly burning LPG burner. The gas cooler after the secondary combustion furnace cools the exhaust combustion gas with cooling water to protect the equipment on the downstream side. However, in this paper the combustible gases were not burned using LPG rather they were diluted and discharged through the stack. However, the gases were monitored for the emissions.

The temperatures at various locations in the CFB rig were continuously monitored and recorded using Yokogawa Hybrid Recorder (HR 1300). The furnace pressure, supply air flow rate were also monitored. The concentrations of emission gases CO, CO₂ and O₂ contained in the produced

gas were also monitored and recorded using calibrated Madur Flue Gas Analyzer (GA-40 plus). The resolution for measurement of O₂ and CO₂ was 0.01 %. The resolution for the measurement of CO and NO was 1 ppm.

2.2 Experimental Method

Approximately 200 gm of the circulating alumina powder was loaded into the experimental CFB and a circulating fluidized bed combustion was formed by supplying air and increasing the in-furnace temperature. Formation of the circulating fluidized bed was observed by monitoring the change in the temperature of the loop seal. It takes about 3-4 hours before the CFB attains the presets temperatures. The presets temperatures were selected using the thermogravimetric analysis (TGA) data showing the devolatilization behaviour. After the CFBC temperature reached the specified temperatures, constant feeding of biomass can be started, and the in-furnace temperature and the gas composition (CO, CO₂ and O₂) needs to be monitored.

After the temperatures stabilized, the primary air flow rate of the fluidizing air was varied and produced gas was analyzed from the sampling port after the cyclone and as well as from exhaust. The biomass feed rate varied depending on the biomass used. The experimental control parameters are being summarized in Table 3.

Table 3: Experimental condition in the CFB test rig

Palm shell powder/blends feed rate		3.0-7.0 kg/hr
Air flow rates	Primary air	1.0 – 3.5 m ³ /hr
	Secondary air	1.0-1.5 m ³ /hr
	Loop seal air	0.1-0.75 m ³ /hr
	Combustion air	10-28 m ³ /hr
Gas Velocity		1.0-2.0 m/s
Gas Residence Time		1.1 – 2.0 sec
In Furnace Temperature		700-850 °C

3. Results and Discussion

3.1 Combustion of Palm shell waste

The oil palm shell waste was obtained from Kulai Palm Oil Mills of Federal Land Development Authority (FELDA), Johor. They were crushed, grind and sieved to have different particle size samples. The combustion experiment was started with palm shell waste particles of 212-300 microns. However, it was found that the palm shell particles were too small and they were easily pushed

by the screw feeder even at a low rpm. This caused a heavy flow of fine palm shell particles in the furnace. The temperature at the top of the CFB was increased rapidly as the feeding started and reached to about 709 °C in a very short time. In order to evaluate the effect of primary air flow through the CFB, the flow rate was varied and the system was stabilized before data was recorded. The emission and experimental data is being shown in Table 4.

Table 4: Emission data for variation in primary air

Sr. No.	Primary Air Flow rate m ³ /hr	Exhaust Temperature °C	CO ppm	NOx ppm	CFB Top Temperature °C
1.	1.0	185	2950	75	709
2.	1.5	160	4819	49	685
3.	2.0	150	7000	38	680

Due to increased particle flow rate, the circulating particle forms dense bed. Palm shell particles (PSP) fed to this layer get into contact with the high-temperature circulating particle and are quickly combust. When palm shell particles (PSP) are in contact with a circulating particle at a higher temperature, the heat conductance to the PSP improves, and the PSP may be more quickly burned. As shown in Table 4, as the primary air flow is increased, the CO concentration increased rapidly. This is probably due to very fine particle size and low residence time in the CFB furnace. This resulted in poor combustion and a high

concentration of CO was found near the CFB top. The emission behaviour showed heavy white smoke representing incomplete combustion.

Keeping in view of the inadequacy of the screw feeder to control the feed rate of the 212-300 microns particles, it was decided to feed bigger PSP of 425 and 600 microns particles. It was found that the 425 microns particles showed adequate feeding behaviour. A typical heating behaviour for top of the CFB rig for 425 microns particles is shown in Figure 3.

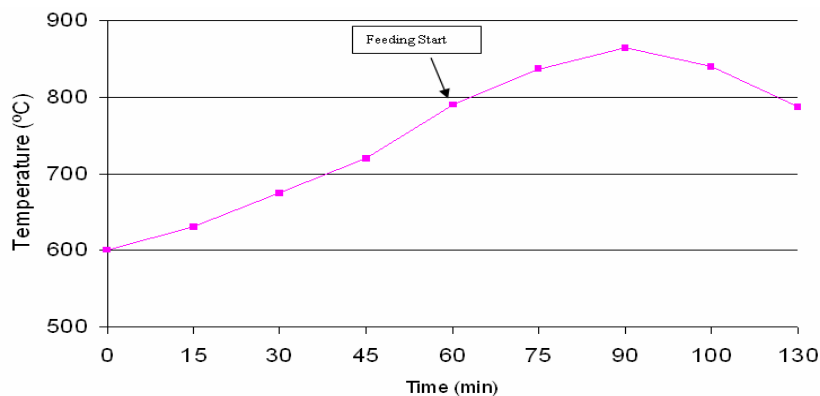


Figure 3: Variation of CFBC temperature with time

The feeding was started at about 60 minutes of stabilization time when the temperature was at 790 °C. As soon as the feeding started, the devolatilisation process

started and temperature of the bed started to rise rapidly. In about 30 minutes time the temperature at the top of the furnace rose to about 864 °C. The effect of the variation of primary air on emissions is being shown in Table 5.

Table 5: Emission data for variation in primary air

Sr. No.	Primary Air Flow rate m ³ /hr	Exhaust Temperature °C	CO ppm	NOx Ppm	CFB Top Temp °C
1.	1.0	105	1094	38	752
2.	1.5	193	6092	127	837
3.	2.0	150	4420	90	864
4.	2.5	143	8500	52	842
5.0	3.0	150	5354	40	809

It can be inferred from Table 5 that the primary air flow rate of 1 m³/hr is good for 425 microns particles. Increasing the flow rate above that reduces the particle residence and hence resulted in incomplete combustion and hence high CO emissions. Similar trend were also shown by the 600 microns particles fluidization. However, in that case, the temperature at the top of the CFB rose to about 888 °C and the CO emissions were about 7000 ppm.

It is important to understand that as the primary air flow rate is changed then the fluidisation behaviour in the combustion also changes abruptly which causes devolatilisation behaviour to change. This can be shown by plotting the average CFB temperature with the height of CFB test rig as shown in Figure 4. This shows a sudden jump in temperature after middle section of the CFB and

it is expected to change the combustion and emission behaviour. It is understood that the gas residence time in the furnace calculated based on the air feed rate is short, 1.1 to 2.0 sec. The residence time may be shorter when the gas flow rate is increased by combustion of PSP. Particularly, under high-temperature conditions, the residence time may be extremely short, less than 1 sec. It is assumed that although the sample was burned by the progress of the quick heat decomposition, that is a feature of the CFB, the reaction of the generated hydrocarbon did not advance because sufficient resident time could not be obtained, and soot and char were generated in quantities.

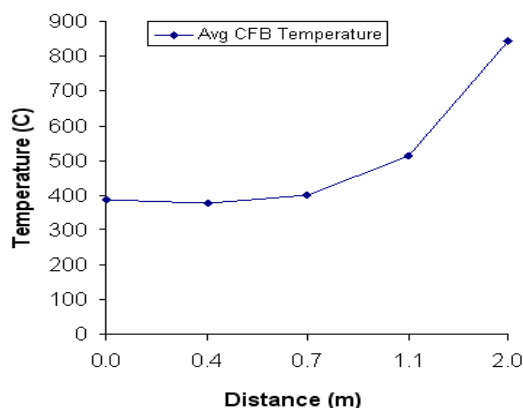


Figure 4: Variation of CFBC average temperature with gasifier height

Soot and char were also found in the duct of the experimental apparatus. This proves insufficient residence time under high-temperature conditions

In a circulating fluidized bed furnace, heat decomposition of the sample supplied into the furnace may advance quickly. In addition, since the particle concentration in the furnace is high and the efficiency of contact of the circulating particle with the sample and gas is high, high heat conductivity and resultant high-speed progress of reaction can be expected. These features of the circulating fluidized bed are regarded as one of the factors of low tar generation. In this experiment, for the particle sizes 212-300 and 425 microns of the supplied PSP were small enough so that the quick heat decomposition of the particles may have been accelerated. The palm shell powders used in this experiment as biomass may have been characterized by easy heat-decomposition and low tar generation. However, slightly higher tar generation was found for the palm shell powder of 600 microns. Nevertheless, reduction in tar generation is always one of major problems for biomass combustion systems. However, our experimental work suggests that good heat transfer between the fluidising particles and palm shell waste can be achieved by having the appropriate amount of alumina particles in the CFBC and using the low primary air flow rates. In order to lower the tar formation further investigations were done for co-firing of palm shell waste with coal blends.

3.2 Co-combustion of coal blends

The tests were conducted for Mukah coal blends ranging from 10-60%. The combustion behaviour was similar to as it was with oil palm shell waste. However, the SO_2 emission was more pronounced as the percentage of coal blending was increased. Figure 5 shows the emission behavior of CO and NOx for various coal blends and it can be inferred that by increasing the blending ratio the CO and NOx emissions can be minimized. However, Figure 6 shows that the SO_2 levels increases sharply as the blending

ratio increasing which may ultimately set the limit for coal blending ratio to be used in pilot CFB combustors.

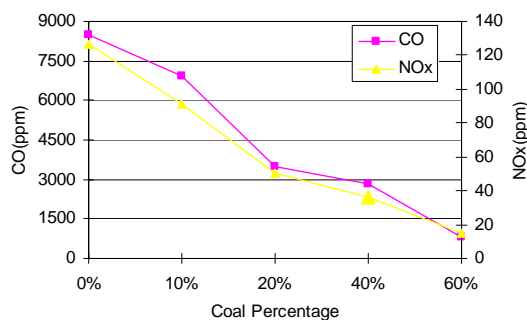


Figure 5: Emission profile for various coal blend

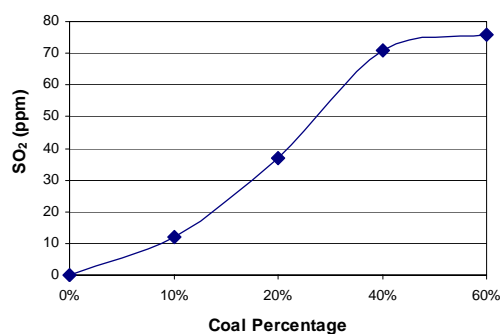


Figure 6: Variation of SO_2 as a function of blend percentage.

4. Conclusions

The combustion experiment were successfully conducted using the circulating fluidized bed. As the result of combustion of oil palm shell waste in the circulating fluidized bed, flue gas containing CO and NOx was produced. In general the tar generation was low as the heat decomposition in the circulating fluidized bed furnace advanced quickly. The dependence of in-furnace temperature and primary air flow on emission behaviour were identified. By this research, we have demonstrated the combustion process of palm shell waste and Mukah coal blends, both of which are found in abundance in Malaysia. Also fundamental data on combustion of coal blends of Mukah coal with oil palm shell waste has been obtained from the CFB

Acknowledgements

The author is deeply grateful to his PhD supervisor, Prof. Dr. Farid Nasir Ani, for his valuable suggestions and guidance in the preparation of this manuscript. The author would also like thank Malaysian Government Commonwealth Secretariat, Public Service Department Malaysia, Putrajaya for funding the PhD program.

References

1. Ani, F.N. (1992) Thermal characteristics of oil palm fruit shells. Proceedings of the Second International energy Conference- Energy from Biomass Residues, Kuala Lumpur, Malaysia.
2. Bhattacharya S.C. (1997) State of the art of Biomass gasification. Proceeding of the International Symposium on Advances in Alternative and Renewable energy, Organized by Universiti Teknologi Malaysia, Johor Bharu, Malaysia.
3. Bridgwater A.V. (1994) Advances in Thermochemical Biomass Conversion, Volume 1, Blakie Academic & Professional, New York.
4. Bui T, Loof R, Bhattacharya SC (1994) Multi-stage reactor for thermal gasification of wood. Energy 19(4):397-404.
5. Cox.R., Chen, S.P., Hon, V., Unya, A. and Batoi, R. (2000) Economic geology and mineral potential of west Sarawak, Malaysia". Proceedings of the Asia Pacific Mining and Quarrying (APMQ) 2000, Kuala Lumpur, Malaysia , pp. 155-174.
6. Devi L., Ptasinski K.J and Janssen F.JJ (2003) A review of the primary measures for tar elimination in biomass gasification processes, Biomass & Bioenergy (24):125-140.
7. Dornburg, V. and Faaij, A. (2001) Efficiency and economy of wood fired biomass energy systems in relation to scale regarding heat and power generation using combustion and gasification technologies, Biomass & Bioenergy 21, (2):91-108.
8. Jamil M. K. and Ani F.N. (1999)Renewable Sources from Pyrolysis of Solid Wastes. Proceedings of the Malaysian Science & Technology Congress, Sarawak, Malaysia.
9. Kate R., Deshmukh G. and Tandale M.S (1997) Energy From Biomass – A perspective under Indian Conditions. Proceeding of the International Symposium on Advances in Alternative and Renewable energy, Organized by Universiti Teknologi Malaysia, Johor Bharu, Malaysia.
10. Pierre Joulain (2004) The efficient combustion of biomass and its derivatives: A major challenge for the future. Proceedings of the Joint International Conference on Sustainable Energy and Environment (SEE), Thailand.
11. PORIM (2005) A summary of the performance of the Malaysian Oil Palm Industry – A weekly update. Palm Oil Research Institute of Malaysia (http://econ.mpob.gov.my/economy/summaryindustry_jan06.htm).
12. Rogner HH, Popescu A. (2000) An introduction to energy. World Energy assessment-Energy and challenge of sustainability New York; United Nations Development Programme.

Value Management in Construction Industry

Bachan Singh^{1*}, Abdul Rahim Abdul Hamid¹, Ng Kim Lai¹

¹ Faculty of Civil Engineering

Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia

Tel: +60-7-5531538, Fax: +60-7-5566157, E-mail: bachan19@yahoo.com

Abstract

The Malaysian government has realized that it is necessary to modernize the construction industry to toe in line with international practice and long term healthy growth. As a result, the use of Value Management as an advanced management technology and skills was highlighted. From the sample interviews carried out randomly, it was found that the participants in the construction industry lack awareness in Value Management. Therefore a study was carried out to examine the typical Value Management job plan methodology, the degree of understanding of Value Management, the main factors that hinder the application of Value Management in the construction industry, and the factors that encourage the implementation of Value Management in the construction industry. The study was conducted in Kuala Lumpur, Johor Bahru and Penang. The study was carried out through interviews and questionnaires. The data analysis was carried out using average index. From the study, the respondents' degree of understanding of Value Management was average. The main factors that hinder the application of Value Management in the construction industry are lack of knowledge about Value Management and lack of support from project owners. While the factors that encourage the implementation of Value Management are the elimination of unnecessary costs and to achieve optimum value for money in satisfying the client's needs.

Keywords: Value Management, Value Management job plan, Lack Awareness, Unnecessary Costs, Optimum Value

1. Introduction

Value management was highlighted with the establishment of Institute of Value Management Malaysia (IVMM) on 9th. October 2001. Value management is currently at the initial stage and has high potential to developed, according to the chairman of the Institute of Value Management, Mazlan Che Mat. Value Management can be said as a proactive, creative problem-solving tool which is a strategic approach to achieving maximum value in a project consistent with the organization's broad business goals. It is a structured team approach to problem solving and reducing costs that can be applied during the objective setting, concept, design and construction stages and the on-going management of buildings [9]. A value management exercise aims to achieve optimum value by providing the necessary functions at the least cost without affecting the specified quality and performance. The benefits of applying value management are to achieve better value for money in satisfying the customer's need, savings in project costs by elimination of unnecessary costs, better understanding of the project's objectives, enhance the function of the project,

* Corresponding Author. E-mail: bachan19@yahoo.com, Tel: +60-7-5531538, Fax: +60-7-5566157

improved team-working among the construction professional, and enhanced the creativity through interaction of different professionals and external experts in construction field.

1.1. Statement of the Problem

The Malaysian Construction Industry contributed about 3.2% of its GDP totaling RM209.27 billion. The possible impact of Value Management implementation in the construction industry can significant as a 10% potential savings in the industry could generate an estimated financial savings up to RM670 million [8].

Value Management has been widely practiced in countries like the USA, United Kingdom, Australia and Hong Kong. Even there are some evidences of Value Management applications in the Malaysian construction industry; but it is still not so popular in Malaysia and lack awareness of its existence and applications. Value Management can be considered still at its infant stage in Malaysia as only a small amount of construction projects have been known to apply Value Management so far [8]. Less than 10% of construction firm had practiced value management to reduce cost due to lack of knowledge. Most of the local developers and government are fresh to its concept [6]. Majority of the contractors in Malaysia are not in favour of Value Management. Contractors prefer to cut

cost without considering the quality, functionality and safety of the structure that is to be built.

Unnecessary costs will exist in both private and public construction firms due to many factors such as lack of information, lack of ideas, honest but wrong beliefs, habits and attitudes, lack of communication and coordination, and out dated or unsuitable standards, codes and specifications. Without the implementation of Value Management studies, the unnecessary costs may not be identified and removed. Consequently, it will result in the cost of the construction project to be high.

1.2. Objectives of Study

1. To study the typical Value Management job plan methodology for construction industry
2. To identify the degree of understanding of Value Management by the construction's professional.
3. To identify the main factors that hinders the application of value management in construction industry.
4. To identify the main factors that encourages the implementation of value management in construction industry.

1.3. Scope of Study

The study focus on construction firms in Penang, Kuala Lumpur and Johor. The respondents of the survey questionnaires include the developers, project managers, civil and structural engineers, quantity surveyors, contractors, Value Management Consultants and Public Works Department personnel.

2. Literature Review

2.1. Value Management

Value is defined as a cost effective way to reliably perform a function that will meet the client's needs, wishes and expectations [1]. Value management can be said as a proactive, creative problem-solving tool which is a strategic approach to achieving maximum value in a project consistent with the organization's broad business goals.

Value Management encompasses the following characteristics [4]:

- It uses a functional oriented.
- It uses teamwork to merging different experiences and skills.
- It is life cycle oriented which examines the total cost of owning and operating a facility.
- It follows the systematic and organized job plan and eliminates the unnecessary costs in a project.
- It directs efforts towards maximum possible alternatives through creativity techniques.

2.2. Unnecessary Cost

Mudge (1971) states that an unnecessary cost is the costs, which do not meaningfully contribute to the product to which they build up. Crum (1971) defined unnecessary cost as the costs that adds nothing to the value of a product, or are not essential to achieve the specified functions.

2.3. Timing of Application

Value Management should preferably be implemented at the early stage of a project's life cycle. This is so because the earlier in the design process that the study is undertaken, the higher the cost reduction potential and the lower the unnecessary design costs. Therefore, delay in the application of value management study until the construction phase would limit its potential of cost reduction.

2.4. Composition of the Value Management Team

Basically, a Value Management team includes a trained, experienced and independent Value Management facilitator from within or outside of the organization; the client; the project team and including end-users.

2.5. Value Management Job Plan Methodology

The Value Management job plan will covers 3 major periods of activities including Pre-Study, Value Management Study and the Post-Study [3].

2.5.1. Pre-Study

The preparation tasks which involve the following six areas such as:

- ❖ Identify client's needs
- ❖ Gathering a complete data file of the project
- ❖ Determining evaluation factors
- ❖ Scoping the specific study
- ❖ Building appropriate models and;
- ❖ Determining the team composition.

2.5.2. Value Management Study / Value-Study

The Value Management session used a typical Value Management job plan methodology, comprising of the following 6 common stages introduced by Pro-Value Management Malaysia:

i. Information

The aim of this stage of Value Management is to provide the information to establish the areas for Value Management study. Its provide opportunity for the participants to understand the issues and constraints, as well as to exchange relevant information among themselves.

ii. Analysis

Function Analysis is basically a method to meet the needs and required functions. Function Analysis System Technique (FAST) diagram was used for analysis of each function to eliminate unnecessary processes and provides a framework to seek better ideas or options.

iii. Creativity / Speculation

This stage involved the generation of ideas in a free and creative way through brainstorming without any criticism. Such technique forces the ideas generation which were focused on the basic functions established at earlier stage.

iv. Judgment / Evaluation

The Judgment or Evaluation stage involved the use of logical thinking to screening and ranking of ideas generated during the Creative / Speculative stage that give the most value improvement potential. The ideas are evaluated by listing the advantages and disadvantages of each idea.

v. Development

This stage usually involves the checking and verification of practicability, fit for purpose, benefits, depending on the selected ideas or alternatives. Then these ideas are expanded into workable solution.

vi. Recommendation

This stage comprises of two phases i.e. first is presentation followed by a written report submission to the top management of the organization for review and reflects upon the Value Management recommendations.

2.5.3. Post-Study

Value Management team should take the responsibility to implement the Value Management change. Progress should be reviewed periodically to make sure that the roadblocks which arise can be overcome.

3. Methods

3.1. General Approach of the Research

Figure 1 shows the methodology flowchart of the study being carried out.

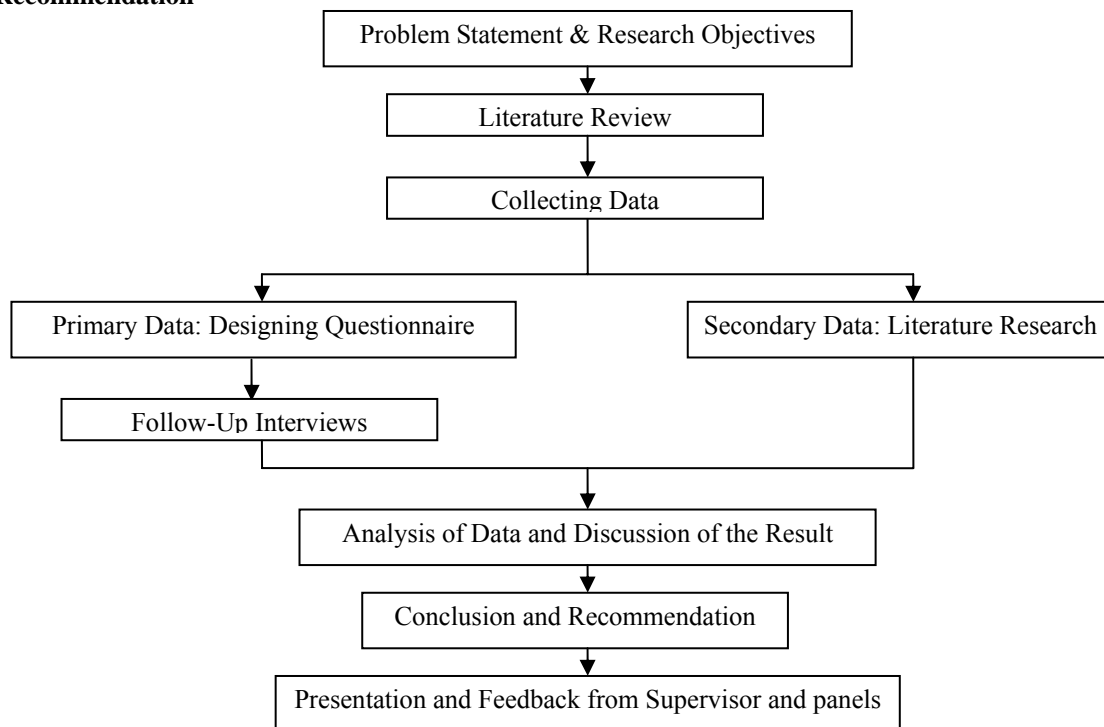


Figure 1: Research Methodology Flow Chart

4. Results and Discussion

To determine the ranking of the various factors captured in the questionnaires, the average index computation was adopted.

$$\text{Average Index (AI)} = \frac{\sum a_i x_i}{\sum x_i}$$

Where:

a_i = The index of a class = 1,2,3,4 and 5 respectively
 x_i = Frequency of the i response

An example of the analysis and result is listed in table 1 as follows:

	Factors that hinders the implementation of Value Management	N	Least Critical	Slightly Critical	Average	Critical	Very Critical	Average Index
a	Lack of time to implement	24	1	4	5	9	5	3.54
b	Lack of knowledge about Value Management	24	0	0	5	12	7	4.08
c	Lack confident to apply it	24	0	0	15	8	1	3.42
d	Lack of support from project owners.	24	1	1	1	13	8	4.08
e	Defensive attitude of the original design team	24	1	1	11	10	1	3.38
g	Interruption to normal work schedule	24	1	2	7	10	4	3.58
i	Difficulty of conducting evaluation and analysis	24	1	2	7	13	1	3.46
j	Too expensive to carry out Value Management	24	0	6	12	6	0	3.00
k	Lack of local VM implementation guideline	24	0	3	7	11	3	3.58
l	Inapplicability of VM principles in Malaysia's construction industry	24	2	2	4	14	2	3.50

Table 1. Factors that hinders the implementation of Value Management

The average index as shown in table 1 for 'lack of knowledge' and 'lack of support' has an average index of 4.08. The average index for 'lack of local VM implementation' and 'interruption to normal work schedule' is 3.58. In terms of ranking, the factors that hinders the implementation of Value Management in terms of priority are : lack of knowledge, lack of support, lack of local VM implementation and so on according to the score of the average index.

5. Conclusion

Based on the study carried out as above, the following conclusions are made:

5.1. The Typical Value Management Job Plan Methodology for Construction Industry

From the study, it was found that the typical Value Management job plan methodology is as follows:

1. Pre-Study
2. Value Management Study
 - Information
 - Analysis
 - Creativity / Speculation
 - Judgment / Evaluation
 - Development
 - Recommendation

3. Post Study

5.2. Degree of Understanding of Value Management by the Construction's Professional.

From the study, it was found that the respondent's degree of understanding about Value Management was 'average'.

5.3. Main Factors that Hinders the Application of Value Management in the Construction Industry

From the study, it was found that the major factors that hinder the application of Value Management in construction industry base on priority are as follows:

- i. Lack of knowledge about Value Management
- ii. Lack of support from project owners
- iii. Interruption to normal work schedule
- iv. Lack of local Value Management implementation guideline
- v. Lack of time to implement
- vi. Inapplicability of VM principles in Malaysia's construction industry
- vii. Difficulty of conducting evaluation and analysis
- viii. Lack confident to apply it
- ix. Defensive attitude of the original design team
- x. Too expensive to carry out Value Management

5.4. The main factors that encourages the implementation of value management in construction industry

From the study, it was found that the main factors that encourage the implementation of Value Management in construction industry are as follows:

- i. Elimination of unnecessary costs
- ii. To achieve optimum value for money in satisfying the customer's needs
- iii. Better understanding of the project's objective
- iv. To cut project cost without adversely affecting the quality and performance or the project
- v. To enhance function of a project
- vi. To arrive at a more effective design
- vii. Identifies project constraints or problems
- viii. Improved team-working
- ix. To encourage creative thinking
- x. Existing of Value Management manual

In order to minimize all the unnecessary costs that exist in the construction project, value management should be applied. Related organization such as IEM and CIDB should establish more seminars or training programs to introduce Value Management in order to increase its implementation in the construction industry.

References

- [1] Dell' Isola, A.J., 1988. *Value Engineering in The Construction Industry*. Smith Hinchman & Grylls, Washington DC Pg 8, 52-68
- [2] Hasnan Abdullah, 1994. *Pengurusan Penilaian Dan Praktis Ukur Bahan*. Unit Penerbitan Akademik, UTM, Skudai. Page 4, 19.
- [3] John W. Bryant, 1998. *Value Methodology Standard*. Society of American Value Engineers. Page 1- 8.
- [4] Kelly JR & Male SP., 1991. *The Practice of Value Management: Enhancing Value or Cutting Cost*. The Royal Institution of Chartered Surveyor, United Kingdom.
- [5] Mazlan Che Mat, 1999. *Value Management-Principal and Application*. Professional Center of Value Management. Pages 38-39.
- [6] Mohd. Zainiddin, 2000. *Keberkesanan dan Penerimaan Pengurusan Nilai di di Dalam Industri Binaan Malaysia*. Projek Sarjana Muda Ukur Bahan.
- [7] Mudge A.E., 1971. *Value Engineering, A Systematic Approach*. McGraw- Hill Book Company, New York. Page 5, 7.
- [8] Ong Hock Teck, 2003 . *Quality and Value Management In Construction Industry*. Pro-Value Management Malaysia. Page 6 – 10.
- [9] SAVE, 2006. Citing Internet sources. URL <http://www.value-eng.org>

Digital Watermarking, Challenges and Applications

Akram M. Zeki¹ and Azizah A. Manaf²

¹ Faculty of Computer Science and Information System
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Tel: +60-19-6340683, E-mail: akramzeki@yahoo.com

² Akademi Tentara Malaysia
Universiti Teknologi Malaysia, ATMA, 57000, Kuala Lumpur, Malaysia
Tel: +60-3-9051-2535, Fax: +60-3-9057-4291, E-mail: azizah@atma.gov.my

Abstract

In the image watermarking context the term visibility or transparency use to designate this concept. Another important requirement is the capacity of the communication channel. The challenge is to embed as much information as possible while staying compatible with the image noise model. The last important requirement about robustness is that it must be possible to detect/remove the hidden message. Obviously, there is a trade-off between the robustness and the capacity of the watermark. In this study different levels of LSB have been used for embedding to test the embedding capacity and PSNR, at this time different quality of JPEG have been applied to study the effectiveness of JPEG on the images.

Keywords:

Watermarking, Watermarking Requirements, Image Quality, Capacity, Robustness.

1. Introduction

The purpose of the watermark is to supply some additional information about the image (image logo, serial number, text message ..) to identify the image owner or a particular customer, to verify image integrity, or to achieve control over the copy process of a particular digital media. The most important requirement in the system we require that the modified object be indistinguishable from the original. In the image watermarking context the term visibility or transparency use to designate this concept. Another important requirement is the capacity of the communication channel. The challenge is to embed as much information as possible while staying compatible with the image noise model. The last important requirement about robustness is that it must be possible to detect/remove the hidden message. This means that the watermark should be readable from images that underwent common image processing operations, such as filtering, lossy compression, noise adding, histogram manipulation, and various geometrical transformations.

The number of bits carried by the watermark could be as low as one bit or several hundred bits. Obviously, there is a trade-off between the robustness and the capacity of the watermark. Another important attribute of watermarking is the computational complexity of the embedding and extracting process. In some applications, it is important that the embedding process be as fast and simple as possible (watermarking images in digital cameras for tamper detection) while the extraction can be more time consuming.

In other applications, the speed of extraction is absolutely crucial (e.g., extracting captions from digital video).

2. Types of Watermarking

Image watermarking can be classified as visible or invisible. A visible watermarking typically contains a visual message or a company logo indicating ownership of the image. An invisible watermarked image is visually very similar but not necessarily identical to the original unmarked image. The invisible watermark's existence should be determined only through a watermark extraction or detection algorithm [1]. There are few applications for each of these types as following: [2]

- **VISIBLE WATERMARK:** Visible watermarks can be used in following cases: Visible watermarking for enhanced copyright protection, and Visible watermarking used to indicate ownership originals.
- **INVISIBLE ROBUST WATERMARKS:** Invisible robust watermarks find application in following cases: Invisible Watermarking to detect misappropriated images, and Invisible Watermarking as evidence of ownership.
- **INVISIBLE FRAGILE WATERMARKS:** Following are the applications of invisible fragile watermarks: Invisible Watermarking for a trustworthy camera. And Invisible Watermarking to detect alternation of images stored in a digital library.

3. Applications of Watermarking

Watermarks added to digital content serve a variety of purposes. The following list details some purposes of digital watermarking [3] & [4]:

- Data Hiding - Watermarks can also be used to hide secret private messages.
- Ownership Assertion – to establish ownership of the content (i.e. image)
- Fingerprinting – to avoid unauthorized duplication and distribution of publicly available multimedia content
- Authentication and integrity verification – the authenticator is inseparably bound to the content whereby the author has a unique key associated with the content and can verify integrity of that content by extracting the watermark
- Content labeling – bits embedded into the data that gives further information about the content such as a graphic image with time and place information.
- Usage control – added to limit the number of copies created whereas the watermarks are modified by the hardware and at some point would not create any more copies (i.e. DVD).
- Content protection – content stamped with a visible watermark that is very difficult to remove so that it can be publicly and freely distributed.

Unfortunately, there is not any universal watermarking technique to satisfy all of these purposes [3]. The content in the environment that it will be used determines the digital watermarking technique.

3. Watermarking Requirements

With respect to the general information hiding problem, a trade off is involved between robustness, visibility and capacity as illustrated in Figure 1.

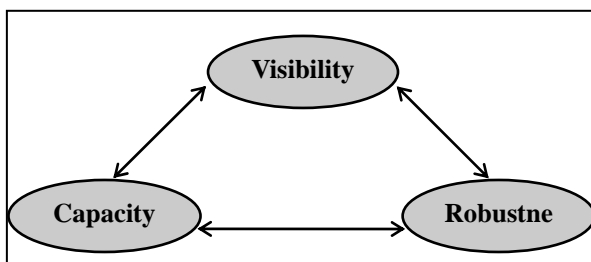


Figure 1. Watermarking Components

3.1. Image quality

In most applications, the watermarking algorithm must embed the watermark such that this does not affect the quality of the underlying host data. The watermark is truly imperceptible if humans cannot distinguish the host data from the watermarked data. However, since users of watermarked data normally do not have access to the host data, they cannot perform this comparison. Therefore, it is sufficient that the modifications in the watermarked data go unnoticed as long as the data are not compared with the original data [5].

Perceptual coders minimize the error perceived by the human visual system (HVS). These were introduced since it was found that working with the peak signal to noise ratio (PSNR) criterion and the mean square error (MSE) criteria was inadequate in reducing perceived distortions introduced by compression [5]. A common measure for compression performance is the achieved compression ratio as shown in equation 1, and the mean squared error MSE can be presented as equation 2.

$$CR = \frac{\# \text{ bits of the original image}}{\# \text{ bits of the compressed image}} \quad (1)$$

$$MSE = \frac{1}{N} \sum (F'_i - F_i)^2, \quad (2)$$

which is the averaged term-by-term difference between the input signal (the original image, F) and the output signal (the watermarked image, F'), the signal to noise ratio as shown in equation 3 and 4.

$$SNR = \frac{\frac{1}{N} \sum_i F_i^2}{MSE}, \quad (3)$$

$$SNR(dB) = 10 \log_{10} SNR, \quad (4)$$

which represents the size of the error relative to the input signal – alternatively on a logarithmic scale, in unit of decibels – or the peak signal to noise ratio (PSNR), given by equation 5.

$$PSNR(dB) = 10 \log_{10} \frac{F_{peak}^2}{MSE}, \quad (5)$$

where F_{peak} is the peak value of the input signal (usually 255 for 8 bit grey scale images).

3.2. Robustness

A second important requirement of watermarking schemes is robustness. Clearly a watermark is only useful if it is resistant to typical image processing operations as well as to malicious attacks. However, it is important to note that the level of robustness required varies with respect to the application at hand.

3.3. Watermarking Capacity

Finally capacity refers to the amount of information we are able to insert into the image. Designing and optimizing information hiding algorithms involves the delicate process of judiciously trading off between these three conflicting requirements [5]. Under the present day scenario a rough estimate of low, medium and high payload, particularly for images, is shown in Table 1.

Table 1. Payload categorization based on message size

Message Size % of cover message	Embedding Capacity
0 – 2 %	Low
2 – 10 %	Medium
10 – 20 %	High
> 20 %	Very High

4. Least Significant Bit Modification

In general, watermark can be embedded in spatial domain or transform domain, [6] in the transform domain approach, some sort of transforms is applied to the original image first. The transform applied may be Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT), etc. The watermark is embedded by modifying the transform domain coefficients of an image.

In the spatial domain approach which is used here, the pixel value of an image is modified to embed watermark information. The spatial techniques insert the watermark in the underused least significant bits of the image. This allows a watermark to be inserted in an image without affecting the value of the image [7]. This technique give the extraordinarily high channel capacity of using the entire cover for transmission in this method, a smaller object may be embedded multiple times. Even if most of these are lost due to attacks, a single surviving watermark would be considered a success.

LSB substitution however despite its simplicity brings a host of drawbacks. Although it may survive transformations such as cropping, any addition of noise or lossy compression is likely to defeat the watermark. An even better attack would be to simply set the LSB bits of each pixel to one...fully defeating the watermark with negligible impact on the cover object. Furthermore, once the algorithm is discovered, the embedded watermark could be easily modified by an intermediate party [8].

The image has been used in this study is gray scale level image (image from University Technology Malaysia) as shown in Figure 2, each pixel in a gray level image is represented by an 8-bit value, where the left bits represents the most significant bit plane and the right bits represents the least significant bit plane.



Figure 2. Gray scale image has been used for testing

4.1. Least Significant Bit and Capacity

To test the ability of LSB technique, different numbers of bits have been used for embedding in order to see the capacity and PSNR for each embedding process. One bit for each pixel has been used first to hide data, then two bits, three bits, four bits and five bits for each pixel, as shown in Table 2.

Table 2. Number of bits have been used for each pixel

No. of bits have been used for embedding	The locations of bits have been used for embedding
1 bit	
2 bits	
3 bits	
4 bits	
5 bits	
6 bits	

The embedding capacity and PSNR value have been calculated for each embedding process for each level of LSB, as shown in Table 3.

Table 3. Embedding capacity and PSNR value for each level of LSB

No. of bits have been used for embedding	PSNR after embedding within selected bits	Embedding capacity for selected bits %
1 bit	60.6764	0.39
2 bits	52.1843	1.17
3 bits	45.2538	2.74
4 bits	38.7265	5.88
5 bits	32.6305	12.15
6 bits	26.5846	24.70

The results show that the maximum number of bits can be used for embedding are 5 bits (last 5 bits) otherwise the watermark will be visible because the PSNR value will be less than 30.

4.1. Least Significant Bit and Robustness

One of the most important attack can be applied into the watermarking images very easy is JPEG or lossy compression, this attack is simple attack that dose not change the geometry of the image and dose not make any use of prior information about the watermark, this attack dose not treat the watermark as noise, three different quality of JPEG have been applied as follows: JPEG90, JPEG70, and JPEG50, simple comparison between the original image and JPEG image to study the effectiveness of JPEG on the images.

The comparison has been done for each level of LSB and the number of bits with positive comparison has been calculated for each quality of JPEG 90, 70 and 50 as shown in Table 5, Table 6, and Table 7.

Table 5. No. of bits can be extracted for each level of LSB with JPEG 90

LSB	No. of bits can be extracted % JPEG 90
1 st	99.6600
2 nd	99.0400
3 rd	97.8433
4 th	95.7917
5 th	91.6200
6 th	83.2767
7 th	70.9050
8 th	61.8925

Table 6. No. of bits can be extracted for each level of LSB with JPEG 70

LSB	No. of bits can be extracted % JPEG 70
1 st	98.8717
2 nd	96.7850
3 rd	92.6533
4 th	85.7508
5 th	73.4717
6 th	61.5592
7 th	54.5392
8 th	51.1458

Table 7. No. of bits can be extracted for each level of LSB with JPEG 50

LSB	No. of bits can be extracted % JPEG 50
1 st	98.7367
2 nd	96.2450
3 rd	91.0925
4 th	82.7783
5 th	69.4483
6 th	58.3150
7 th	52.4292
8 th	50.1017

The results show that the effectiveness of JPEG attacks increase with small LSB.

5. Conclusion

With respect to the general information hiding problem, a trade off is involved between robustness, visibility and capacity. The host image which has been tested here is gray scale image (8 bits per pixels) where as the watermark object has been embedded at different levels of LSB with the aim to study and calculate the embedding capacity and PSNR value. The effectiveness of JPEG attack has been studied also. Simple comparison between the original image and JPEG image has been done for each level of LSB and the number of bits with positive comparison has been calculated for each quality of JPEG 90, 70 and 50, the results show that the effectiveness of JPEG attacks increase with small LSB.

Acknowledgments

I would like to thank University Technology Malaysia, for sponsoring this research under IRPA grant number: 7424904.

References

- [1] Yang, Z. Dual Domain Semi-fragile Watermarking for Image Authentication, Master thesis, University of Toronto, Canada 2003.
- [2] Yeung, M.M et al. "Digital Watermarking for High-Quality Imaging", 1997 IEEE First Workshop on Multimedia Signal Processing, Jun 23-25, 1997, Princeton, New Jersey, pp 357-362.
- [3] Memon, N., & Wong, P. W. (1998). Protecting digital media content. Communications of the ACM, 41(7), 35-43.
- [4] Jian Zhao, Eckhard Koch, and Chenghui Luo_ In business today and tomorrow Communications of the ACM . 41. (7) 67-72 July 1998.
- [5] Shelby P. Robust Digital Image Watermarking, PhD thesis, University of Geneva, Faculty of Science, Canada. 2000.
- [6] F.A.P. Petitcolas, R.J. Anderson, M.G. Kuhn, "Information Hiding-A survey", Proceedings of the IEEE, Volume:87 Issue: 7, Page(s): 1062-1078. 1999.
- [7] M. George, J.Y. Chouinard and N.D. Georganas, "Spread Spectrum Spatial and Spectral Watermarking of for Images and Video", Proc. 1999 IEEE Can. Workshop in Information Theory (CWIT'99), Kingston, Ont., June 1999.
- [8] C. Shoemaker, "Hidden Bits: Techniques for Digital Watermarking", Technical report, Chris Independent Study. 2002.

Measurement of Health Risk Events using RISKAS v2 for Arsenic Contamination

Sri Adelila Sari^{1*}, **Zaini Ujang**², **Umi Kalthom Ahmad**¹, **Lukman Hakim Ismail**²

¹Department of Chemistry, Faculty of Science

²Institute of Environmental and Water Resource Management

Universiti Teknologi Malaysia (UTM), 81310 Skudai, Johor, Malaysia.

Tel: +607-5531578, Fax: +607-5531575, *adelila@yahoo.com

Abstract

This study aimed at measuring health risk events using a tool developed for environmental health risk assessment, known as RISKAS v2. This tool is developed as measure and manages health risk events in sequential decision. A variety of alternative and complementary measures of the health risk events were examined using RISKAS v2 in sequential decision, represented as single or multiple objective decision trees. An example demonstrates the incorporation of different measures of risk events in a multi-objective decision tree for arsenic contamination in Tasik Biru, Bau, Sarawak. The results showed that measures of the health risk events could be interpreted, thereby enabling to measures of the risk events to be handled in straightforward manner in event tree.

Keywords:

Health risk, event tree, RISKAS v2, arsenic contamination

1. Introduction

The toxicological properties of arsenic have been known for centuries (William, 2001). Symptoms of chronic arsenic poisoning have been recorded in populations reliant on water supplies containing more than 50 µg/L arsenic in several countries, including Chile, Mexico, Taiwan, India and Thailand (William, 2001). This value currently constitutes the permissible limit of the European Union (EU), the United States Environmental Protection Agency (USEPA) and many other countries with respect to arsenic in potable water. Epidemiological evidence of adverse effects at lower exposure level has, however, prompted the World Health Organization (WHO) to promote an interim guideline of 10 µg/L (WHO, 2004).

At Tai Parit, Sarawak, an open-cast gold mining pit has been operated by the Borneo Company since 1989. After being flooded in 1921, it was rehabilitated and redeveloped as a tourist spot known as Tasik Biru. However, Tasik Biru has been declared unsafe for recreational and tourism activities in 1999 due to high level of arsenic in water (New Straits Time, 1999). The Natural Resources and Environmental Board (NREB) as the regulator for environmental quality in Sarawak reported a concentration of 0.87 parts per million (ppm) of arsenic in the lake water in 1999 (New Straits Time, 1999) and increased up to 2.75 ppm in 2001 (New Straits Time, 2001), which is above the 0.05 ppm (50 µg/L) of drinking water quality stated in the guidelines issued by the Malaysia Ministry of Health (Tan, 1999).

Due to the possibility of residents living close to

post-mining lakes being exposed to arsenic, therefore health risk assessment need to be emphasized to ensure the effective health protection and improvements. It is also to support the framework of health-based targets for safe drinking water established by WHO, (2004). The intention of this study was to display a useful framework for decision analysis of uncertainties about the relationship of arsenic to human cancer.

In the case of arsenic, there has been relatively little application of information concept in decision analysis to regulation of toxic chemical. It perhaps because of the difficulties involved in placing monetary values on human health effects, such as human cancer. When decisions involve uncertainties, more information can be valuable. Decision analysis should offer a method of calculating the value of obtaining new information. The method is based on observation that information is valuable to the extent that it help us make better decisions, so we can value information by comparing a decision before and after hypothetical information is available.

One of graphic tool for decision analysis is the decision tree. This tool is designed to highlight the ramifications of alternative decisions and uncertain events. Like a tree, it continues to branch out as it grows. As a tool for gaining insight into a problem, a decision tree can be used both qualitatively and quantitatively, just like an influence diagram. For a quantitative analysis, the tree can be solved to identify the optimal decisions for a particular problem.

* Corresponding Author. E-mail: adelila@yahoo.com,
Tel: +60-7-5531578, Fax: +60-7-5531575

2. Approach and Methods

2.1. RISKAS v2 in perspective

RISKAS v2 is the decision support system (DSS) that developed to assist users to assess the risks of polluted chemical substance to people. The conventional computer programming language has been selected in the development was Microsoft Visual Basic 6.0. In addition, Microsoft Access 2002 was used for database management. The system is performed on the PC-Windows environment. Figure 1 shows the front page of RISKAS v2.



Figure 1: The front page of RISKAS v2

2.2. Event Tree in RISKAS v2

The building blocks of a event tree in RISKAS v2 were based on two types of nodes: a *decision* represented by the first horizontal line and a *chance* represented by success and failure square. At a decision node, the branch which is needed to follow must be chosen. At a chance node, there is uncertainty about the outcome of a future event. Each branch from the node represents one of several possible outcomes covering the complete possibilities. A probability which usually expressed as a fraction was assigned to each of these possible outcomes. Figure 2 shows the 3 “event tree” in RISKAS v2.

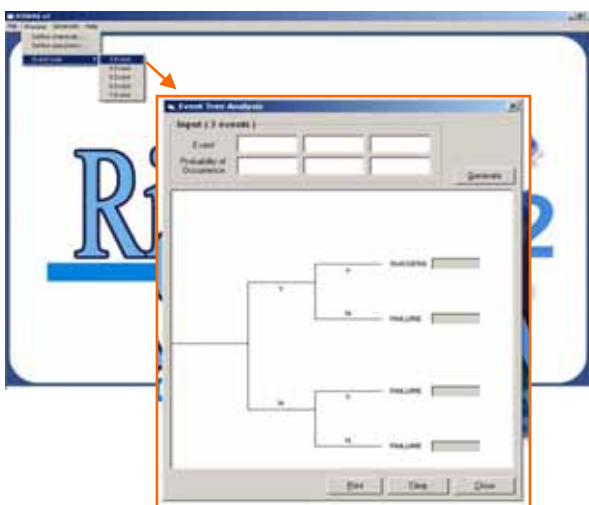


Figure 2: Event tree pop up box in RISKAS v2

2.3. Description of Sampling Area

Bau is one of the twenty-eight districts in Sarawak. It is situated in the Western part of the state of Sarawak in Malaysia and is linked by approximately 35 km of winding road to the state capital, Kuching. It has an area of 884.40 square kilometers sharing a common border with Kalimantan Barat, Indonesia. The district capital is Bau town, which known as the gold town of Sarawak, due to its rich gold ore deposits and gold-mining activities in the past. Sampling was performed at nine points in Tasik Biru, Bau, Sarawak as shown in Figure 3. Sampling points, Pt 2 to Pt 4 were situated near the edge of the lake, approximate one meter inward. In addition, Pt 5 to Pt 9 were situated in the mid part of the lake, with Pt 9 at a point very close to the adjoining inlet of Sungai Bau and Pt 1 was located close to the adjoining outlet of Sungai Bau.

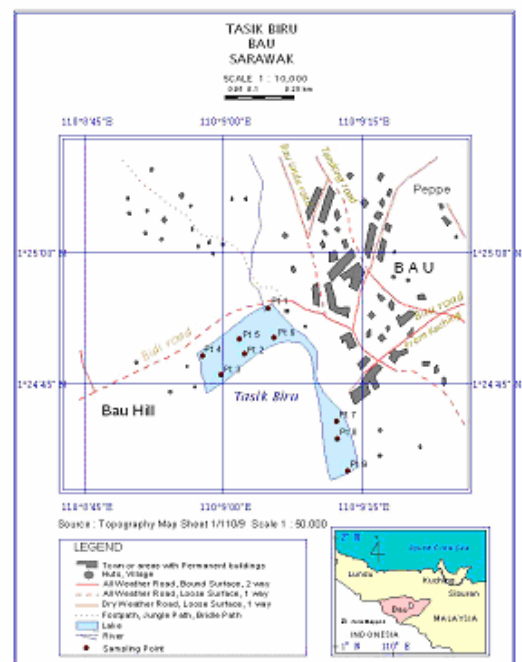


Figure 3: Sampling map

2.4. Sample Preparation

Samples of the lake water were performed at 9 sampling sites and taken every 5 meters deep to a maximum depth of 70 meters and immediately preserved with concentrated nitric acid (HNO₃) to pH less than 2. Samples were taken every six months with four different sampling periods, started from March 2003 to September 2004. The depth were measured using Bathy 500 MF echo sounder, meanwhile the position were acquired using Garmin 12XL (Handled GPS) and also HYDROpro navigation as a hydrographic software to manage all data. Total arsenic, iron, and manganese were measured using an Inductively Coupled Plasma-Mass Spectrophotometer (ICP-MS) ELAN 6000 (Perkin-Elmer SCIEX, USA) instrument and the ionic species were determined using an HP^{3D} Capillary Electrophoresis (CE) (Agilent Technologies, Waldbronn, Germany).

3. Results and Discussion

3.1. Arsenic in Tasik Biru Lake

In Tasik Biru lake, the average concentrations of arsenic were recorded from 0.704 to 0.927 mg/L, whereas the maximum was 2.730 mg/L on August 2003. It increased from 0.527 mg/L in epilimnion zone (0 to 10 m) to approximately 2.730 mg/L in hypolimnion zone (45 to 70 m). In addition, the variations of arsenic species were also found to be similar as the total arsenic. It can be seen in Figure 4, whereas concentration of As (V) and As (III) were in the range of 1.560 mg/L to 1.280 mg/L, respectively.

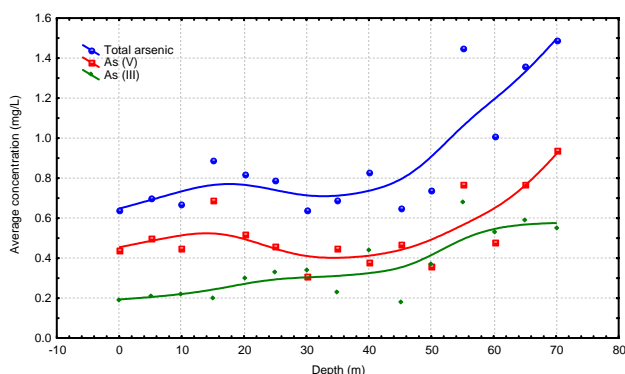


Figure 4: Depth profiles of arsenic in Tasik Biru

The concentration of arsenic in Tasik Biru lake exceeded the maximum permissible level of arsenic in drinking water of 0.05 mg/L for Malaysia and 0.01 mg/L for WHO standard values. Based on arsenic speciation, arsenic in the form of As (III) also exceeds the standard of drinking water. Therefore, the concentration of arsenic in Tasik Biru lake water was dangerous to human health.

3.2. Health risk event for ingested inorganic arsenic

The decision on health effects of arsenic in drinking water could be structured with help of an event tree. An event here defined relative to a frequency of occurrence, i.e. events with a severity of consequences that is only infrequently reached or exceeded [6].

The event tree was used to three different of arsenic concentration in Tasik Biru Bau, Sarawak. The example began by constructing an event tree reflecting the arsenic concentration and probability of occurrence decision and important uncertainties as depicted in Figure 5.

As shown in Figure 5, the uncertainty of risk probability is represented as a single horizontal line of 1.0 mg/L arsenic concentration with two branches. Each branch corresponds to alternative concentrations. The first branch represents the possibility of arsenic concentration of 0.62 mg/L and the second branch represents the possibility of arsenic concentration of 0.5 mg/L.

The probability risks of exposure to ingested arsenic in drinking water was predicted using RISKAS v2, furthermore the values were input to event tree analysis in RISKAS v2.

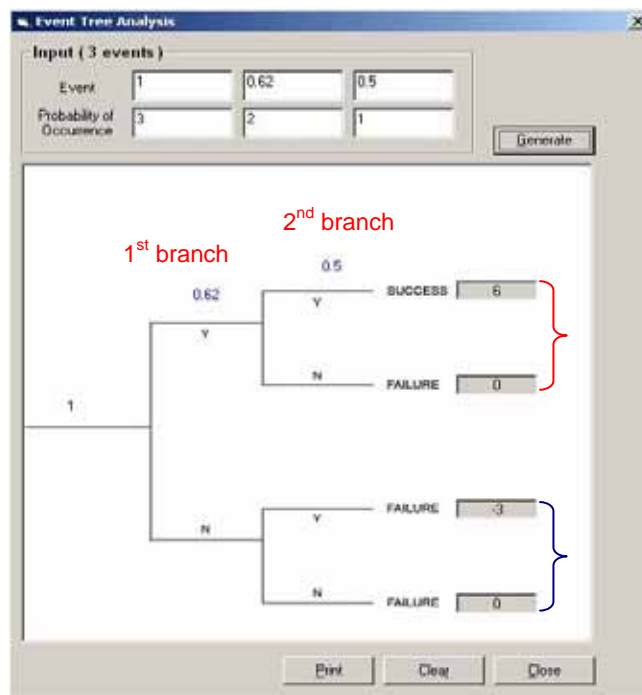


Figure 5: Event tree depicting choice among arsenic concentration of 1.0, 0.62 and 0.5 with probability of arsenic’s risk. All values are illustrative.

The concentrations of arsenic were input as event and the probabilities of risk were input as probability of occurrence by assuming: $3 = a$, $2 = b$ and $1 = c$.

Results of event tree calculation as shown in Figure 5 could be defined as follows:

$$\begin{aligned}
 6 &= a \times b \times c & (1) \\
 0 &= (a \times b) \times (1-c) & (2) \\
 -3 &= a \times (1-b) \times c & (3) \\
 0 &= a \times (1-b) \times (1-c) & (4)
 \end{aligned}$$

The result values are predicted values, that is, probability of population exposed to risk of ingested water contaminated with arsenic. There were 6 chances that effected to arsenic exposure and 3 chances that people will not effected. The illustrative analysis suggests that chance of people exposed to arsenic were found to be higher. The analysis based on risk calculation using RISKAS v2 reinforced this observation, whereas the probability of risk will be higher on higher level of arsenic concentration and vice versa.

Value-of-information risk assessment and event tree offer the alternative to risk management tool. Since the risk assessment information could be used for determining clean-up levels for hazardous sites, such as Tasik Biru, the analysis uses event tree as illustrative data could be also used for cost analysis data. For example, the hazardous sites with higher level concentration of arsenic will more costly than lower level to be clean up.

4. Conclusion

Arsenic concentration in Tasik Biru lake was found to be exceeded the maximum permissible level of arsenic in drinking water. From this point of view, the Tasik Biru water must not considered healthy for human consumption. The water contains high concentrations of arsenic, which would have adverse effects on human health if it is consumed on a daily basis.

The frequency occurrence of risk of exposure to ingested arsenic could be analyzed using event tree. With the help of event tree, the value-of-information to estimate the potential chance could be considered. Furthermore, it could be as alternative to manage the risk.

the risk management could be structured

Acknowledgments

Thanks are due to the Ministry of Science, Technology and Environment for financial support (IRPA RM 8 vot 74028)

References

- [1] New Straits Time (1999). "Tasik Biru Unsafe due to Arsenic in Water. 28 October.
- [2] New Straits Time (2001). "Finding Ways to Reduce Arsenic in Lake Water". 26 May.
- [3] WHO. (2004). "Guidelines for Drinking – Water Quality". 3rd (Ed.). Geneva.
- [4] William, M. (2001). "Arsenic in Mine Waters : an International Study". *Environ Geol.* 40 (3) : 267-278.
- [5] Tan, H. (1999). "Challenges in Meeting National Demand for Safe Water". *Bul Kes Masy.* Isu Khas. 21-34.
- [6] Frohwein, H. I., Lambert, J. H. and Haimes, Y. Y. (1999). "Alternative Measures of Risk of Extreme Events in Decision Trees". *Rehab Eng & Syst Safety.* 66 : 69-84.

Vibration Characteristics of Shape Memory Alloy Hybrid Composite Plate

W.C. Tan, A. Murni, M. I. Ghazali and S. Jamian

Faculty of Mechanical and Manufacturing Engineering
Kolej Universiti Teknologi Tun Hussien Onn
Beg Berkunci 101, 86400 Parit Raja, Batu Pahat, Johor Darul Takzim, MALAYSIA
Email: tweechoon@hotmail.com

Abstract

The orientation angles of the Shape memory alloys (SMAs) had been identified as an important parameter that will affect the performance of the hybrid composite. This paper studies on the effect of different orientation of SMAs wire which were embedded into hybrid composite plate. The specimens were in the form of epoxy/fiber glass reinforcement that embedded with SMAs wire (Muscle Wire – Flexinol 375 HT) to form a polymer matrix composite. Different angles of orientations of SMAs wire will be embedded to the hybrid plate. The vibration tests on the specimens were done by using a shaker that received the wave signal from signal generator and amplified by the power amplifier. The vibrations from the samples were measured by Polytech model PDV 100 predictive maintenance equipment (laser Doppler vibrometer). The results showed that with different angles of orientation of the SMAs wire, it shifted the fundamental frequency and affect the vibration characteristic of the hybrid composite plate.

Keywords:

Shape Memory Alloy, hybrid plate composite, vibration

1. Introduction

Shape memory alloys (SMAs) are widely used in the engineering applications such as the sensor, actuator and damping instruments. The hybrid composites that embedded with SMAs showed some unique properties or functions such as self strengthening, active modal modification, high damping, damage resistance, control and self healing. They can provide tremendous potential in many engineering applications. Since vibration control is one of the most important and common applications of the shape memory alloys, the study of vibration characteristic of SMAs become important. Past researcher like Jae-Sang Park *et al.* [1] studied on the vibration of thermally post-buckled composite plates embedded with shape memory alloy fibers. Although the shape memory alloys had been manufactured and used in the modern smart structure, but the major internal problem of shape memory alloy still exist as a result of the characteristic of shape memory that limited to motion along the fibers' longitudinal axes [2]. The inter-relationship between the angle of SMAs embedded into the composite and the effect of different angle of SMAs embedded into the composite to the vibration characteristic have not been studied and determined by previous researchers.

For the vibration control, Lammering and Schmidt [3] addressed that the SMAs is an ideal material that improved the damping properties of structures, which required to load dynamically within the transformation range from austenite to martensite. Beside this, Travis L. Turner [4] had

experimentally validated the model that have been developed for predicting thermo-mechanical behavior of the shape memory alloy hybrid composite structures with an embedded SMAs constituent. In his research, the static and dynamic experimental configurations of shape memory alloy hybrid composite specimens were described and experimental results for thermal post-buckling and random response were presented. Excellent agreement was achieved between the measured and predicted results, fully validating the theoretical model on constrained recovery behavior of shape memory alloy hybrid composite structures. Kaori Yuse and Yoshihiro Kikushima [5] purposed a safety SMAs actuator from which no debonding occurs even when the maximum shear stress was generated by recovery force in SMAs wire. Kaori Yuse and Yoshihiro Kikushima demonstrated that the control effect on beam structure was around 5 Hz. The results clearly showed that SMAs device can be in an active control loop with an enlarged frequency bandwidth. The SMAs/CFRP hybrid actuator generates much higher force and displacement compared to other actuator devices. Xu and Song [6] proposed shape memory alloy wall joint to adaptively attenuate and control the vibration wave propagation in cylindrical shells. This wall joint inserted along the cylindrical shells acted as a source of mismatch of structural impedance with tunable characteristic and having the potential to solve the pass-band problem. The numerical results demonstrated that the adaptive SMAs wall joint with a suitable temperature able to control the vibration from a source with wideband frequencies or time-varying frequencies, and the transmission loss was more than 20dB. Rustighi *et al.* [7] then developed a real time control of a shape memory alloy adaptive tuned vibration absorber

whereas heating or cooling the shape memory alloy changed its elastic modulus and hence the effective stiffness and tuned frequency of the adaptive tuned vibration absorber. The result showed that it is possible to use SMAs to realize and control a continuous adaptive tuned vibration absorber and that industrial application is therefore feasible. Marek Pietrzakowski [8] studied the modification of dynamic properties of rectangular laminated plates and sandwich plates containing layers reinforced with shape memory alloy fibers. In his work, the influence of parameters of “heating” and “cooling” processes on the system response has been analyzed. The numerical results showed that the range of natural frequency changes depends on the SMAs fiber volume fraction in activated layers and can be reduced by activation of only a part of the SMAs fibers.

In this paper, there were four specimens with different angle orientations of the arrangement of SMAs in the composite to be fabricated. These four specimens are the composite plate that was without SMAs fabricated as a reference for analysis, and for the sample which embedded with Muscle Wire type Flexinol 375 HT with diameter 0.375mm were oriented at 0° , 45° and 90° to the epoxy/fiber glass reinforcement composite plate. The specimens were tested on the boundary conditions of clamped-free. The frequency of the wave received from the laser vibrometer for the composite plate will be studied.

2. Specimens

The specimens with the dimension of 200 mm x 150 mm x 2 mm was fabricated from the epoxy resin and reinforced with the type E glass fiber using the hand lay-up technique that similar to the technique that stated by W.C. Tan *et. al.* [9]. The SMAs that used in the testing were the Flexinol wire from Mondo Tronics, Inc. Detail parameter of Flexinol wire can refer to www.Mondo.com. The volume fraction of the Flexinol wire has been designed to be 0.05% and the Flexinol wire layer was located in the mid-plane of the composite plate. The Flexinol wire will be activated at 90°C .

3. Experiment setup

For the Flexinol wire, two types of testing conditions were done; the inactivated condition and the activated condition. The term of activated refer to the Flexinol wire in which the current will flow through the wire and heated up until the phase transformation occur. The temperature measured on the outer surface of the Flexinol wire was allowed to stabilize for few minutes before the measurement is recorded with a thermocouple. The layout of experiment is shown in Figure 1. The shaker which was modified from speaker excite at the frequency of 1 kHz. The input signal of the speaker was measured by the digital oscilloscope to confirm the signal that generated by the signal generator to the speaker.

For the experiment setup, the electrical heating must be done very careful so that the temperature of the Flexinol wire will not exceed the temperature that can damage the properties of the Flexinol wire. The jig to hold the specimen must be careful design and checked so that the jig does not vibrate with the specimen during testing. The locations where the vibration readings were taken are shown in Figure 2.

Similar experiment setup was done to determine the fundamental frequency of the specimens. The vibration signal was then measured at point 21 with the laser vibrometer.

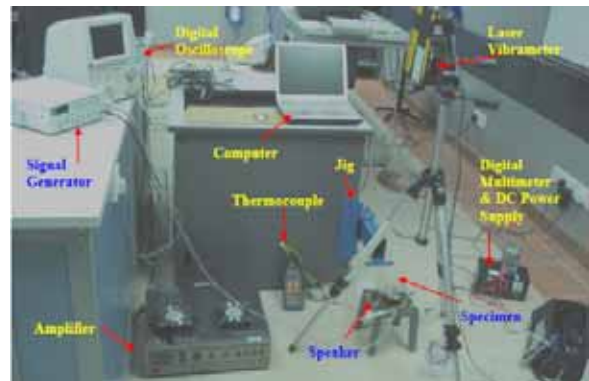


Figure 1. Experiment setup for the vibration test.



Figure 2. Locations of points for the specimen.

4. Data and analysis

The vibration test was conducted on the clamped-free for all four samples. The locations of points where the reading were taken are shown in Figure 2. The direct currents that applied to the Flexinol wire were 0.50A, 0.75A, 1.00A and 1.25A when excited by the forcing frequency but for the experiment to determine the fundamental frequency were 0.60A and 1.20A.

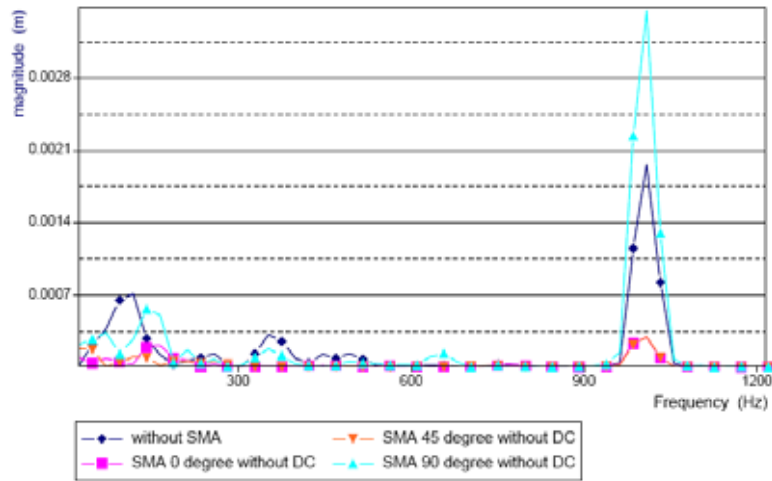


Figure 3. Magnitude-frequency analysis for specimens without DC power supply at point 23

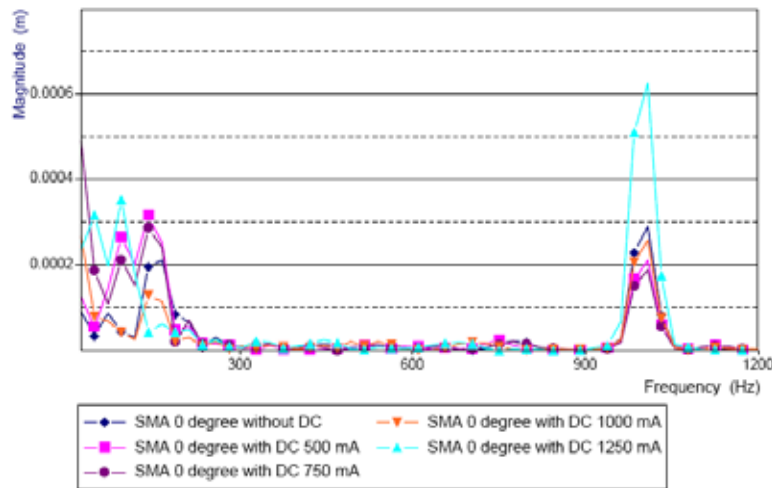


Figure 4. Magnitude-frequency analysis for specimen embedded 0 degree SMA at point 23.

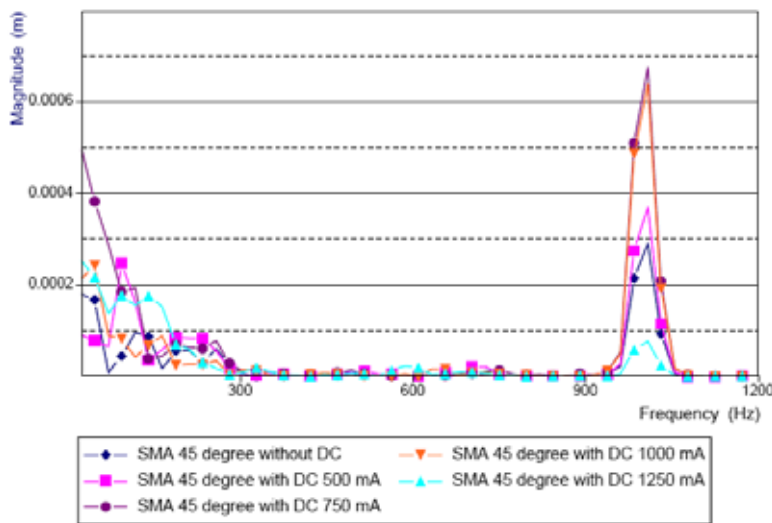


Figure 5. Magnitude-frequency analysis for specimen embedded 45 degree SMA at point 23.

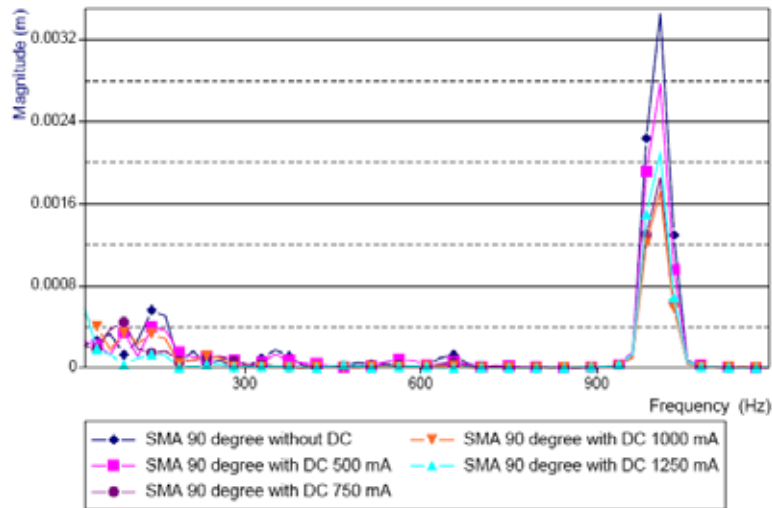


Figure 6. Magnitude-frequency analysis for specimen embedded 90 degree SMA at point 23.

The fundamental frequency that measured from the specimens are given as in Table 1. It showed that when the temperature increase or the changed of phases within the SMA will shift the fundamental frequency of the structure itself and reduced the amplitude of the fundamental frequency of the composite plate.

Table 1. Fundamental frequency of specimens

Specimen		Fundamental Frequency (Hz)	Frequency Shifted (%)
Without SMA		20.508	-
SMA 0 Degree	DC = 0 A	24.902	-
	DC = 0.6 A	24.902	0.00
	DC = 1.2 A	23.438	-5.88
SMA 45 Degree	DC = 0 A	24.902	-
	DC = 0.6 A	20.508	-17.65
	DC = 1.2 A	20.508	-17.65
SMA 90 Degree	DC = 0 A	21.973	-
	DC = 0.6 A	21.973	0.00
	DC = 1.2 A	19.043	-13.33

Figure 3 till Figure 6 showed the result from the experiment where the composite plates were excited at 1kHz with the clamped-free boundary condition. Figure 3 showed the result of the specimens where the vibration signal was measure without apply any current to increase the temperature of the SMA which embedded within the composite plate.

While Figure 4, Figure 5 and Figure 6 showed the result for the specimens which the SMA wire oriented into 0 degree, 45 degree and 90 degree to the longitude axes of the composite plat during different value DC currents were applied onto the SMA wire and heated up the SMA. When the DC current heat up the SMA wire, phase transformation occurred and caused the fundamental frequency of the

composite plate to reduce as shown in Table 1. All measurement were taken at the point 23 where according to W.C. Tan *et. al.* [9] the center of the composite plate having the greater potential to reduce the amplitude of the vibration and this mean that the center point is the most effective point to control the vibration amplitude.

Table 2. Amplitude at resonance frequency

Specimens	DC (A)	Amplitude at 1kHz (m)	Amplitude difference (%)
Without SMA	-	0.001958	-
SMA 0 Degree	0.00	0.000290	-
	0.50	0.000211	-27.16
	0.75	0.000188	-35.18
	1.00	0.000258	-11.08
	1.25	0.000625	115.11
SMA 45 Degree	0.00	0.000291	-
	0.50	0.000369	26.82
	0.75	0.000674	131.38
	1.00	0.000639	119.27
	1.25	0.000078	-73.07
SMA 90 Degree	0.00	0.003450	-
	0.50	0.002770	-19.70
	0.75	0.001853	-46.29
	1.00	0.001714	-50.32
	1.25	0.002097	-39.22

Table 2 showed the amplitude of the resonance caused by the forcing frequency at 1 kHz which shown in Figure 3, Figure 4, Figure 5 and Figure 6. While Figure 7 showed the amplitude of the specimens at the resonance frequency 1 kHz.

The specimen with 0 degree Flexinol wire oriented

having the amplitude of 290 μ m without any current applied on to the specimen but the amplitude reduced to 211 μ m and 188 μ m when the direct current 0.50A and 0.75A was applied. But then the amplitude increased to 258 μ m and 625 μ m when 1.00A and 1.25A current was applied to the Flexinol wire.

The specimen with 90 degree Flexinol wire oriented also showed a similar result where having the amplitude of 3.45mm without any current applied and the amplitude reduced to 2.77mm, 1.853mm and 1.714mm when 0.50A, 0.75A and 1.00A direct current applied onto the Flexinol wire. When heated with 1.25A, the amplitude of the specimen increased to 2.097mm.

While for specimen with 45 degree Flexinol wire oriented, a different result has been observed where the amplitude without applied current was 291 μ m but the increase to 369 μ m and 674 μ m when heated by 0.50A and 0.75A current. Then decrease to 639 μ m and 78 μ m when heated up to 1.00A and 1.25A.

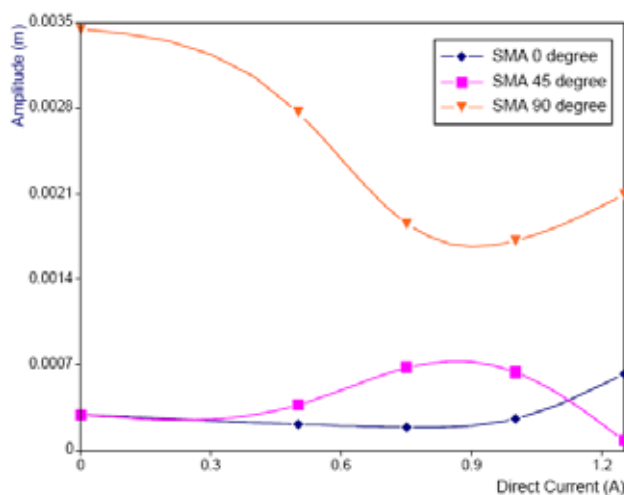


Figure 7. Amplitude at resonance frequency 1 kHz

5. Conclusion

The different angles of orientation of the Flexinol wire on the composite plate will shift the frequency of vibration in both activated and inactivated condition. During inactivated condition, the Flexinol wire behave like a normal reinforcement agent and shift the frequency of the composite plate due to the increasement of total mass within the composite plate and when current flows through Flexinol wire and heated up to the phase transformation temperature, the wire was activated. Thus the shape memory effect of the Flexinol wire will cause the composite plate having inner stress and changed the stiffness of the composite plate and shifted the frequency. The magnitude where the frequency was shifted is varied with the angle of orientation for the SMA fibers within the composite plate. The results also showed that with different angle of orientation for the SMA wire will cause the different vibration characteristic of the composite plate. And based on the results that had been obtained, the SMA wire

oriented at 45 degree was found to be the most effective orientation to shift the fundamental frequency which gave the reduction of 17.65%.

Acknowledgments

The authors would like to thank the Kolej Universiti Teknologi Tun Hussien Onn for sponsoring this work under the grant RMC-0148.

References

- [1] Jae-Sang Park, Ji-Hwan Kim and Seong-Hwan Moon. 2004. "Study on the Vibration of thermally post-buckled composite plates embedded with shape memory alloy fibers." *Composite Structures* 63 179-188.
- [2] Anthony Kelly and Carl Zweben. 2000. "Comprehensive composite materials, Volume 5 – Test methods, Nondestructive evaluation, and Smart Materials." Pergamon, Amsterdam 623.
- [3] Lammering R and Schmidt I. 2001. "Experimental investigations on the damping capacity of NiTi components." *Smart Mater Struct.* 2001 10 853-859
- [4] Travis L. Turner. 2001. Experimental validation of a thermoelastic model for SMA hybrid composites. *SPIE 8th Annual International Symposium on Smart Structures and Materials; Modeling, Signal Processing, and Control in Smart Structures*, SPIE Vol 4326, Paper No. 4326-24.
- [5] Kaori Yuse and Yoshihiro Kikushima. 2005. "Development and experimental consideration of SMA/CFRP actuator for vibration control." *Sensors and Actuators A*: 122 99-107.
- [6] M. B. Xu and G. Song. 2004. "Adaptive control of vibration wave propagation in cylindrical shells using SMA wall joint." *Journal of Sound and Vibration*.278 307-326.
- [7] E. Rustighi, M J Brennan and B R Mace. 2005. "Real time control of a shape memory alloy adaptive tuned vibration absorber." *Smart Materials and Structures*.14 1184-1195.
- [8] Marek Pietrzakowski. 2000. "Natural frequency modification of thermally activated composite plates." *Mec. Ind.* 1 313-320.
- [9] W.C. Tan, A. Murni, M.I. Ghazali and S. Jamian. 2006. "Shape memory alloy hybrid composite plate for vibration control." In *Proceedings of the International Conference on Composite Materials and nano-Structures (iC2MS-06)* Shah Alam, Malaysia.: International Conferences on Composite Material and nano-Structures, Inc.

Optimization of Cellulase and Lignin peroxidase Production by Locally Isolated *Aspergillus* spp. Using Oil Palm Empty Fruit Bunch (OPEFB)

Salwa Mohamad*, Zaharah Ibrahim, Madihah Md Salleh

Biology Department, Faculty of Science,
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
Tel: +60-7-5532528 E-mail: cellewa@yahoo.com

Abstract

Malaysia produces an abundant mass of oil palm wastes every year. Oil palm empty fruit bunches (OPEFB) are disposed by the land filling method, which is very costly, while some are burnt in furnaces, which causes air pollution. The application of OPEFB has been practiced as mulch in oil palm field nurseries and has shown beneficial effects on oil palm growth and yield increases in mature palms. The feasibility of using OPEFB fibres as a substrate for decomposition by *Aspergillus* spp was studied using Mandels' basal medium with OPEFB fibres (1% (w/v)) added as the sole carbon source in the medium. Cultures were incubated in a shaker at 200 rpm, 37°C for 8 days. OPEFB was proven as the most suitable carbon source for degradation compared to oil palm tree trunk, carboxymethylcellulose and cellulose as it produced 2 times higher cellulases (endoglucanase, FPase and β -glucosidase) and lignin peroxidase activities with 2.9, 0.6, 0.01 and 0.06 U/mL, respectively. The highest enzyme activities were observed at 2% (w/v) of OPEFB fibres. pH 5.5 gave the highest specific activities of cellulases (0.35, 0.05, 0.01 U/mg) as compared to pH 4.5, 6.5 and 7.5. *Aspergillus* spp was able to produce high enzyme activities at 37°C as compared to 30°C and 45°C. Peptone gave the highest production of cellulases (endoglucanase, FPase and β -glucosidase) and lignin peroxidase (1.35, 0.18, 0.24, 0.12 U/ml). 0.6% (w/v) of peptone was the most suitable concentration to be used as it enhanced the production of cellulase by 2.3 times as compared to cultures with 0.2, 0.9 and 1.2% (w/v) of peptone in the medium. Optimization of enzyme production by *Aspergillus* spp produced cellulases and lignin peroxidase besides degrading the OPEFB fibres to smaller molecular weight substance such as β -glucose.

Keywords: oil palm empty fruit bunch, cellulase, lignin peroxidase

1. Introduction

Malaysia produces an abundant mass of oil palm wastes every year. In most of palm oil mills in Malaysia, oil palm empty fruit bunches (OPEFB) are disposed by the land filling method, which is very costly [11]. In some factory, the empty fruit bunches are burnt in the furnaces, which causes air pollution.

With the increasing cost of inorganic fertilizers, the use of oil palm by-products as mulch and sources of nutrients is an important alternative. The application of OPEFB has been practiced as mulch in oil palm field nurseries and has shown beneficial effects on oil palm growth and yield increases in mature palms under different soils and rainfall regimes in Malaysia [10].

To fully utilize the potential of OPEFB as biofertilizer, it has to be converted by chemical and/or biological processes. In nature, there are many microorganisms, bacteria as well as fungi that live on natural lignocellulose such as OPEFB. OPEFB may be decomposed on site by filamentous fungi and recycled to

the soil with improved biofertilizer and bioprotective properties [13].

For OPEFB biodegradation, the nature and properties of the extracellular enzyme systems produced by fungi are well characterized. The enzymes include various cellulases, oxidases such as lignin peroxidase, hemicellulases and glycosidases. In this paper, a potential lignocellulose-degrading fungus (*Aspergillus* spp. SN1) has been used throughout the study. The scope of this research was to optimize production of lignocellulose-degrading enzymes by manipulating the physical and environmental factors such as pH, temperature, carbon and nitrogen requirements to obtain the optimal condition for enzyme production.

2. Materials and methods

2.1 Microorganism

Aspergillus spp. strain SN1 used in this study was isolated from sludge samples obtained at a palm oil mill in Johor, Malaysia. The fungus was grown on potato

* Corresponding Author. Email: cellewa@yahoo.com, Tel: +60-7-5532528

dextrose agar (PDA) at 37°C for 8 days for the development of green spores and the spores were harvested. Spore suspension was then stored at 4°C for subsequent use in inoculum preparation.

2.2 Materials

The oil palm empty fruit bunch (OPEFB) obtained from the palm oil mill was shredded into long fibres and dried at 70°C for 2 days. The OPEFB was then cut into small pieces (0.2 – 2.0 cm) for use as substrate.

2.3 Culture medium and shake flask experiment

The basal medium described by Mandels and Weber (1969) contained (g/L): (NH₄)₂SO₄, 1.4, KH₂PO₄, 2.0, Urea, 0.3, CaCl₂, 0.3, MgSO₄, 0.3, FeSO₄, 5.0 x 10⁻³, MnSO₄.H₂O, 1.6 x 10⁻³, ZnSO₄.H₂O, 1.4 x 10⁻³, CoCl₂, 2.0 x 10⁻³, peptone, 0.75, Tween 80, 2.0 mL. OPEFB fibres (1% (w/v)) were added as carbon source in the medium. In all of the shake flask experiments, 100 mL medium (pH 5.5) was transferred into a 500 mL shaking flask and inoculated with 10 mL of the spore suspension containing approximately 4 x 10² spores/mL. The flasks were incubated at 37°C for 8 days with agitation on a rotary orbital shaker. Samples were taken at regular time intervals during the 8 days incubation.

2.4 Analytical procedures

The culture withdrawn during the fermentation process was centrifuged at 4000 rpm for 15 min at 4°C. The clear supernatant was analyzed for soluble protein and enzyme activities. Endoglucanase or CMCase activity was determined by measuring reducing sugar produced when CMC was used as substrate while cellobiohydrolase or FPase activity was determined by estimating the reducing sugar liberated from Whatman filter paper No.1 [15].

Enzyme assays were performed in 0.05 M sodium acetate buffer at pH 5.0 and incubated at 50°C for 30 min for CMCase and 1 h for FPase. Reducing sugar content was determined using the dinitrosalicylic acid method [9]. One unit of CMCase and FPase activity was defined as 1 μmol reducing sugar released/mL enzyme/min. β-glucosidase was determined using the method described by Wood and Bhat (1988). In this method, the *p*-nitrophenol released from *p*-nitrophenyl-β-D-glucopyranoside was measured using a spectrophotometer. One unit of β-glucosidase activity was defined as 1 μmol of *p*-nitrophenol liberated/mL of enzyme/min. Lignin peroxidase (LiP) activity was determined based on the oxidation of the dye azure B [1]. One unit of enzyme activity is expressed as an OD decrease of 0.1 units per minute per mL of the culture filtrate. Specific activity is defined as units per mg of protein. Protein content was determined by the method of Lowry *et.al* (1951) using bovine serum albumin as a standard protein.

3. Results and discussion

3.1 Effect of different types of carbon sources

A comparison of enzyme production using OPEFB fibres and other commercial cellulose is given in Table 1. OPEFB fibre was a better substrate for cellulase and lignin peroxidase (LiP) production compared to carboxymethylcellulose (CMC), oil palm tree trunk or cellulose. Among the four carbon sources tested, OPEFB fibre gave the highest production of CMCase and FPase, followed by cellulose and CMC as shown in Table 1. At 3 days of incubation, the CMCase activity in cultures containing CMC as substrate was only 2% of CMCase activity with OPEFB as a substrate. Cultures containing OPEFB showed activities of FPase about 10 times higher than the cultures containing CMC. The lignin peroxidase activity in medium containing OPEFB was about 1.3 times higher than that obtained in medium containing CMC.

The use of different types of carbon source greatly influenced the pattern of cellulase production. Table 1 showed that CMCase and FPase activities increased with time until it reached the peak after about 4 days of incubation in culture containing OPEFB.

Table 1: Enzyme production and biomass yield by *Aspergillus spp.* using different carbon sources.

Carbon source (g/L)	Maximum activity (U/mL)				Protein (mg/mL)
	CMCase	FPase	β-glucosidase	LiP	
CMC	0.055	0.056	0.003	0.025	1.907
Cellulose	2.171	0.289	0.006	0.015	9.966
OPEFB	2.932	0.581	0.009	0.062	10.518
Oil Palm tree trunk	2.453	0.155	0.009	0.143	6.021
	Specific activity (U/mg)				X _{max} (g/L)
	CMCase	FPase	β-glucosidase	LiP	
CMC	0.029	0.029	0.002	0.013	1.023
Cellulose	0.218	0.029	0.001	0.002	5.836
OPEFB	0.279	0.055	0.001	0.006	9.572
Oil Palm tree trunk	0.407	0.026	0.001	0.024	5.867

The reason for faster hydrolysis of OPEFB was probably that it contained more amorphous cellulose and other easily degraded polymers than cellulose which is in microcrystalline form and therefore more likely to be recalcitrant to attack by cellulase. The degradability of cellulose is generally accepted to be inversely related to its crystallinity [3].

It is well known that the nature of carbon source greatly influences growth and level of cellulase activity of fungal cultures. Variations in enzyme activities are due partly to the changes of pH of the culture (which is strongly influenced by the major carbon source) and quality of the substrate. Generally, growth on OPEFB gave higher cellulase production than on oil palm tree trunk, CMC and cellulose with 9.572 g/L of biomass yield.

3.2 Effect of OPEFB Concentrations

Different concentrations of OPEFB (0.5, 1.0, 1.5 and 2.0% (w/v)) were added to the basal medium. The maximum cellulases activities were achieved when the OPEFB concentration was 2.0% (w/v) as shown in Table 2. This result is different compared to the study by Coutts and Smith (1976) which reported that cellulase activity by *Sporotrichum thermophile* was obtained at 1% (w/v) of solka floc. Cellulase enzymes were in excess at low substrate concentrations, thus the enzyme was accumulated after all binding sites on the OPEFB were saturated. However, in the case of higher substrate concentrations, some enzymes were bound to substrate while some enzyme remained free in the culture medium. Meanwhile, maximum lignin peroxidase activity was achieved when the OPEFB concentration was 1.0% (w/v).

Table 2: Maximum enzyme production and biomass yield by *Aspergillus spp.* at different OPEFB concentrations.

[OPEFB] (%) w/v	Maximum activity (U/mL)				Protein (mg/mL)
	CMCase	FPase	β-glucosidase	LiP	
0.5	0.666	0.133	0.020	0.090	4.659
1.0	1.150	0.161	0.030	0.136	7.161
1.5	1.724	0.243	0.050	0.090	9.776
2.0	2.233	0.327	0.080	0.047	14.055
	Specific activity (U/mg)				X _{max} (g/L)
	CMCase	FPase	β-glucosidase	LiP	
0.5	0.143	0.029	0.004	0.019	6.324
1.0	0.161	0.022	0.004	0.019	9.638
1.5	0.176	0.025	0.005	0.009	13.465
2.0	0.159	0.023	0.005	0.003	15.654

3.3 Effect of different types of nitrogen sources

Among the nitrogen sources investigated, results showed that peptone gave the maximum production of lignin peroxidase and all three components of cellulase complex in *Aspergillus spp.*, followed by yeast extract, (NH₄)₂SO₄ and urea as shown in Table 3. The culture pH increased with growth and became constant at the stationary phase, except when (NH₄)₂SO₄ was used as the nitrogen source. The decrease in culture pH in fermentation using (NH₄)₂SO₄ was due to ammonium ions, which were absorbed by the mycelium and subsequently inhibited the production of cellulase enzymes [15].

It was proven that high C/N ratio in the medium

enhanced the production of enzyme. Meanwhile, protein concentrations were high when peptone and yeast extract were used as nitrogen sources as compared to urea and (NH₄)₂SO₄. High productions of enzymes were detected when organic nitrogen sources (peptone and yeast extract) were used. This might be because yeast extract and peptone contained significant amount of carbon, vitamins and minerals, hence enhancing growth of the fungus and subsequently gave higher enzyme activities [16].

Table 3: Enzyme production by *Aspergillus spp.* using different types of nitrogen sources.

Nitrogen source	Maximum activity (U/mL)				Protein (mg/mL)
	CMCase	FPase	β-glucosidase	LiP	
(NH ₄) ₂ SO ₄	0.259	0.050	0.048	0.071	2.847
Urea	0.142	0.022	0.044	0.058	2.157
Peptone	1.347	0.178	0.245	0.124	10.159
Yeast extract	0.763	0.138	0.244	0.033	6.230
	Specific activity (U/mg)				C:N
	CMCase	FPase	β-glucosidase	LiP	
(NH ₄) ₂ SO ₄	0.091	0.018	0.017	0.025	2.244
Urea	0.066	0.010	0.020	0.027	2.380
Peptone	0.133	0.018	0.024	0.012	4.287
Yeast extract	0.122	0.022	0.039	0.005	4.857

β-glucosidase activities were low when (NH₄)₂SO₄ and urea were used as nitrogen sources compared to peptone and yeast extract. The most important property in the saccharification of cellulose is the ability of cellobiohydrolase to act synergistically with endoglucanase (CMCase) to hydrolyze crystalline cellulose, since neither is effective on its own. During the reaction, cellobiose is produced and becomes inhibitory to both types of cellulase and, hence, β-glucosidase is required to hydrolyse cellobiose to relieve the inhibition of endoglucanase and cellobiohydrolase [16]. The maximum activities of CMCase, FPase, β-glucosidase and lignin peroxidase obtained with peptone as nitrogen source were 1.347, 0.178, 0.243 and 0.124 U/mL, respectively.

3.4 Effect of peptone concentrations

Peptone concentrations ranges from 0.2-1.2% (w/v) were used to study its effects on enzyme production. The optimum concentration of peptone for cellulase and lignin peroxidase production was 0.6% (w/v). From Table 4, it showed that high concentration of peptone gave high concentration of protein, therefore decreasing the specific activities of enzymes. Peptone concentrations above 0.9%

(w/v) slightly inhibited cellulase production. Production of

cellulase is optimum under limitation of nitrogen supply.

At low concentrations of peptone, cellulase and lignin peroxidase activities were high, with high levels of C/N ratio. However, increment of peptone concentrations gave higher production of protein. Microbial decomposition of plant material resulted in an overall loss of cellulose and carbon and increase in nitrogen content [14]. Increased microbial protein and humic substances were also detected during decomposition.

Table 4: Maximum enzyme production by *Aspergillus* spp. using different peptone concentrations.

[Peptone] (%) w/v	Maximum activity (U/mL)				Protein (mg/mL)
	CMCase	FPase	β -glucosidase	LiP	
0.2	2.637	0.049	0.048	0.050	8.507
0.6	2.819	0.069	0.052	0.058	10.690
0.9	1.332	0.047	0.039	0.067	13.070
1.2	1.345	0.060	0.045	0.039	14.150
	Specific activity (U/mg)				C:N
	CMCase	FPase	β -glucosidase	LiP	
0.2	0.310	0.006	0.006	0.006	4.287
0.6	0.264	0.006	0.005	0.005	1.429
0.9	0.102	0.004	0.003	0.005	0.953
1.2	0.095	0.004	0.003	0.003	0.715

3.5 Effect of temperature

Different incubation temperatures were used to study its effects on enzyme production. From Table 5, it can be seen that the amount of cellulase and lignin peroxidase produced was affected by the incubation temperature. Maximum CMCase, FPase, β -glucosidase and lignin peroxidase were found at 37°C. *Aspergillus* spp was capable of producing high cellulase and lignin peroxidase over a range of 30°C to 37°C. At high incubation temperatures, denaturation of extracellular enzyme activities occurs on prolonged incubation. As temperature increases, the rate of cellulase and lignin peroxidase production increases. However, at the highest temperature tested (45°C) enzyme activities were low. The optimum temperature for enzyme production is a reflection of protein synthesis and enzyme formation.

The majority of fungi are mesophiles which grow between 5°C and 37°C, which an optimum temperature of 25-30°C [4]. However, in the compost environment the elevated temperature means that the small group of thermophilic fungi is an important biodegradation agent. In nature, thermophilic fungi grow in garden compost heaps, bird nests, coal tips, power plant cooling pipes and effluents, in the storage of many agricultural products and in piles of wood chips and peat. In mushroom compost, thermophilic fungi are responsible for the degradation of

edible fungus [12].

Table 5: Maximum enzyme production and biomass yield by *Aspergillus* spp. in different temperatures.

Temp. (°C)	Maximum activity (U/mL)				Protein (mg/mL)
	CMCase	FPase	β -glucosidase	LiP	
30	0.037	0.045	0.002	0.022	3.313
37	0.229	0.686	0.009	0.042	5.945
45	0.201	0.120	0.006	0.017	8.257
	Specific activity (U/mg)				X_{max} (g/L)
	CMCase	FPase	β -glucosidase	LiP	
30	0.011	0.014	0.001	0.005	8.504
37	0.039	0.115	0.002	0.001	9.700
45	0.024	0.015	0.001	0.004	9.660

3.6 Effect of pH

Different pH of the medium were used to study its effects on enzyme production. The pH of the culture changed during the growth of *Aspergillus* spp (data not shown). Usually, the pH of the cultures increased during the first days due to the fungus utilizing peptone, hemicellulose and amorphous cellulose from lignocellulosic for growth. However, the pH then decreased from day 4 onwards due to the formation of carboxylic acid from lignin as lignin peroxidase is released. At this stage the fungus started to utilize the crystalline portion of cellulose and produced cellulase. The changes in pH observed occurred due to the weak buffering capacity of the medium which contained only 2 g/L of KH_2PO_4 .

The maximum CMCase, FPase and β -glucosidase activities were found at 8, 6 and 4 days of incubation, respectively as shown in Table 6. CMCase, FPase and β -glucosidase activities were maximum in cultures with an initial pH of 5.5. At pH 6.5, activities were decreased up to 66% and further decreased to 80% at initial pH of 7.5. Meanwhile, maximum lignin peroxidase activity was detected at pH 6.5, with 3 days of incubation.

Most fungi prefer an acidic environment but tolerate to a wide range of pH, with the exception of the Basidiomycotina which do not grow well above pH 7.5. *Coprinus* species are the only Basidiomycotina which prefer an alkaline environment [4]. In addition, not only the enzyme profiles accounts for the capacity of degradation but also their physiochemical characteristics when they are secreted; more stable enzymes at their optimum pH activity are capable to produce a more extensive degradation of the substrate [5]. All of the enzymes had high activities at pH 5.5.

Table 6: Maximum enzyme production and biomass yield by *Aspergillus* spp. in different pH.

lignocellulose, which is a prerequisite for the growth of the

pH	Maximum activity (U/mL)				Protein (mg/mL)
	CMCase	FPase	β -glucosidase	LiP	
4.5	2.018	0.333	0.017	0.036	7.532
5.5	2.742	0.354	0.053	0.119	7.947
6.5	1.412	0.242	0.040	0.098	7.377
7.5	1.000	0.306	0.042	0.079	9.034
	Specific activity (U/mg)				X_{max} (g/L)
	CMCase	FPase	β -glucosidase	LiP	
4.5	0.268	0.044	0.002	0.005	7.212
5.5	0.345	0.045	0.007	0.015	7.992
6.5	0.191	0.033	0.005	0.013	6.972
7.5	0.111	0.034	0.005	0.009	7.488

4.0 Conclusions

Cellulase and lignin peroxidase activities were maximum when OPEFB was used as substrate. *Aspergillus* spp was able to grow and produced high amounts of complete cellulase complex which is essential for rapid hydrolysis of lignocellulosic materials. The maximum activities of all cellulases (endoglucanase, cellobiohydrolase and β -glucosidase) obtained from batch fermentation of OPEFB were about 2 times higher than those obtained from fermentation of cellulose as a substrate. During cellulose hydrolysis, cellobiose is produced and inhibits endoglucanase and cellobiohydrolase; hence high levels of β -glucosidase are required to hydrolyze cellobiose to relieve the inhibition. The optimum initial pH for cellulase and lignin peroxidase production is pH 5.5 with significant increase of pH during the early stage of fermentation. 2.0 % (w/v) of OPEFB fibre in the medium gave the maximum cellulase production while lignin peroxidase was highest produced at 1.0% (w/v) of OPEFB. Different nitrogen sources were used in the medium and peptone gave the maximum CMCase, FPase β -glucosidase and lignin peroxidase activities. Peptone concentration at 0.6% (w/v) in the medium gave the highest cellulase and lignin peroxidase activities. Higher concentrations of peptone (0.9 and 1.2 %) produced high concentration of protein, therefore, limiting the specific activities of these enzymes. Optimization of enzyme production by *Aspergillus* spp produced cellulases and lignin peroxidase besides degrading the OPEFB fibres to smaller molecular weight substance such as β -glucose.

Acknowledgments

I gratefully acknowledge my supervisors, Dr.Zaharah Ibrahim and Dr. Madihah Md. Salleh for kind support and knowledge sharing in this study.

References

- [1] Arora, D.S and Gill, P.K. 2000. Comparison of two assay procedures for lignin peroxidase, *Enzyme and Microbial Technology*, 28: 602-605.
- [2] Coutts, A.D., and Smith, R.E. 1976. Factors Influencing the Production of Cellulases by *Sporotrichum thermophile*, *Applied Environmental Microbiology*. 31: 819-825.
- [3] Cowling, E.B., and Brown, B. 1969. Structural Features of Cellulosic Materials in Relation to Enzymatic hydrolysis, In *Cellulases and Their Applications* (Gould, R.R. ed) American Chemical Society. 152-187.
- [4] Dix, N.J., and Webster, J. 1995. *Fungal Ecology*. Chapman & Hall, Cambridge, Great Britain.
- [5] Lechner, B.E., and Papinutti, V.L. 2006. Production of lignocellulosic enzymes during growth and fruiting of edible fungus *Lentinus tigrinus* on wheat straw. *Process Biochemistry*. 41. 594-598
- [6] Lowry, O.H., Rosebrough, N.J, Farr, A.L., and Randall, R.J., 1951, Protein measurement with the folin phenol reagent. *Journal of Biological Chemistry*. 193: 265-275.
- [7] Ma,A.N., Tajima Y., Asahi, M and Hanif, J.,1996. A Novel Treatment Process for Palm Oil Mill Effluent. *PORIM Technology*. Malaysian Agricultural Research and Development Institute (MARDI).
- [8] Mandels, M and Weber, J. 1969. The production of cellulases. *Advances in Chemistry Series*. 95: 391-414.
- [9] Miller, G.L. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugars. *Analytical Chemistry*. 31: 426-428.
- [10] Ortiz, R.A, Villalobos, E. and Fernandez, O. 1992. Mulch and Fertilizer Effect on Soil Nutrient Content, Water Conservation and Oil Palm Growth, *ASD Oil Palm Papers*.
- [11] Prasertsan, S and Prasertsan, R. 1996. Biomass Residues from Palm Oil Mills in Thailand: An Overview on Quantity and Potential Usage, *Biomass and Bioenergy*.11: 387-395.

- [12] Sharma, H.S.S. 1989. Economic importance of thermophilous fungi. *Appl. Microbial. Biotechnol.* 31.1-10.
- [13] Tengerdy, R.P., and Szakacs, G. 2002. Bioconversion of lignocellulose in solid substrate fermentation, *Biochemical Engineering.* 13: 169-179.
- [14] Wood, D.A, Matcham S.E and Fermor, T.R., 1996. Production and Function of enzymes during lignocellulose degradation'. In Zadrazil F and Reiniger, P. *Treatment of Lignocellulosics with White Rot Fungi.* Elsevier Applied Science Publishers.43
- [15] Wood, T.M, and Bhat, K.M. 1988. Method for measuring cellulase activities. In *Methods in Enzymology. Cellulose and hemicellulose,* Wood, W.A, and Kellog, J.A., (eds) 160: 87-112. Academic Press, New York.
- [16] Zakaria, Z.Z., Khalid H and Hamdan A.B. 2003. Guideline on land application of POME, *PORIM Bulletin.* Palm Oil Research Institute, Malaysia. 28

Pineapple Waste Juice Utilization for Ethanol Production by *Saccharomyces Cerevisiae*

Abdullah, Yefta, Popi

Department of Chemical Engineering, Faculty of Engineering,
Diponegoro University, Semarang. Tel: +62-24-7460058, e-mail: abd_busairi yahoo. com

Abstract

Utilisation of fruits waste is important not only to solve the disposal problem but also to get raw material or for conversion into useful and higher value added products. The pineapple waste juice contains mainly glucose, fructose, sucrose and other nutrients. It therefore can potentially be used as carbon source for alcoholic fermentation. Alcohol is one of the most important organic material is widely used in the beverages industries and the others. The experiments were carried out at temperature of 30 °C and initial pH of 4,0. Effect of various parameters such as inoculum size, nitrogen concentration and initial sugar concentration were studied. The optimal values for tested the variables for batch culture were found to be : sugar concentration, 100 g/l ; nitrogen concentration, 325 mg N/l and inoculum size of 15% v/v with the maximum of alcohol yield of 35.85 %. If the alcohol production using fed-batch culture the yield was increased and obtained the maximum of alcohol yield of 48.27%.

Keywords: Pineapple waste juice, ethanol fermentation, *saccharomyces cerevisia*.

1. Introduction

Major world producers of canned pineapple are Thailand (39%), Philippines (23%), Indonesia (13%), Kenya (8%) and Malaysia (.3%) which all together contribute to more than 80% of total world canned pineapple production.[1-2]. Food processing operation uses enormous quantities of water which are consequently discharge as a polluted effluent. The waste are contain high concentration of biodegradable organic material and suspended solid. As a result it has a high BOD and extremes of pH conditions [3].

The solid waste from pineapple canning process was estimated about 40 - 50 % from fresh fruit as pineapple peals and cores [4]. This waste contains valuable components which are mainly sucrose, glucose, fructose and other nutrients [5]. An attempt has been made by many researchers to utilise the waste for producing high value added chemicals such as single cell protein (SCP), acetic acid, oxalic acid, methane, acetic acid, citric acid and lactic acid [5-8].

Alcohol is one of the most important organic material is widely used in the beverages industries and the others. It therefore can potentially be used as carbon source for alcoholic fermentation.

The objective of this study was to examine the potential of pineapple waste juice as a carbon source for ethanol production by *Saccharomyces cerevisiae* as well as to study the effect of various fermentation parameters such as total sugar concentration, nitrogen concentration and inoculum size.

2. Methods

2.1. Media

The fermentation media used to carry out the fermentation process were solid pineapple waste (peel and core) obtained from local market in Semarang (Indonesia).

2.2. Strain

The microorganism used in this study was *Saccharomyces cerevisiae* ATCC 9763 from Oxoid, UK. It was maintained on potato dextro agar and transferred to fresh medium every month.

2.3. Inoculum

Saccharomyces cerevisiae from Potato Dextrose Agar was transferred to Sabouraud medium (3%), The composition of the medium are Pancreatic digest of casein, 0.5%; Peptic digest of fresh meat, 0.5%; and Glucose, 2%. The cultures were incubated at 30 °C for 24 hours.

2.4. Fermentation

The medium composition of fermentation was varied according to the the experimental design. The medium was then sterilized at 121 °C for 20 minutes, cooled and inoculated. The inoculated flasks were kept on a rotary shaker maintained at 100 rpm and 30 °C for 4 days.

2.5. Analysis

The ethanol concentration were measured using HPLC (Waters TM 600). A 150 mmX7.8 mm ID Phenomenex column with UV detector at maximum wave length of 210 nm were used. The eluent used were sulphuric acid 0.005 N and acetonitril (80:20) at flow rate of 2.0 ml per minute at ambient temperature [9]. The sugar concentration were measured by enzymatic methode using spektrofotometer UV-VIS1700 (Shimadzu) at maximum wave length of 507 nm.

2.6. Experimental design and Optimization

The response surface method was used for the optimization of sugar concentration, nitrogen concentration and inoculum size to the ethanol production.

This technique provides the experimental design and fits the experimental data by using a least squares method to calculate mathematically or identify the optimal response condition of the system [10].

3. Result and Discussion

3.1. The characteristics of pineapple waste juice

The characteristics of pineapple waste juice obtained from this work and published by different authors are given in Table 1. The total sugar concentration is different, it might be due to different process generated wastes, season and area (1. Indonesia, 2. Malaysia and 3. Thailand), but the total nitrogen is quite similar. The pH is higher to those reported by other researcher previously. The pH increases with decreasing total acidity, which indicated by decreasing citric acid content in the pineapple waste juice.

Table 1. The characteristic of pineapple waste juice

Parameter	This Work (2005) ¹	Abdullah (2002) ²	Sasaki (1991) ³
Total sugar (g/l)	91.35	75.47	100.00
Total nitrogen (g/l)	0.14	0.16	0.15
Phosphorus (mg/l)	23	12.0	-
Citric acid (g/l)	1.2	3.84	-
Acidity, as citric acid (g/l)	1.2	3.95	-
pH	4.39	4.0	4.0

3.2. Optimization

The range of variables investigated were : total sugar concentration of 50 and 150 g/l, nitrogen content of 100 and 300 mg/l and inoculums size of 5 and 15% v/v with centre point of 100 g/l, 200 g/l dan 10 v/v, respectively. The ethanol concentration obtained from the experiments carried out according to the experimental design is given in Table 2. By the statistic analysis yielded the following regression equation which is an empirical

relationship between ethanol yield and the test variables in coded unit given in equation (1)

Table 2. 2³ factorial design with 3 centre point

Run	Total sugar (X ₁)	Nitrogen content (X ₂)	Inoculums size (X ₃)	Ethanol
	g/L	mg N/L	% v/v	g/l
1	50	100	5	10,51
2	50	100	15	14,59
3	50	300	5	16,12
4	50	300	15	16,69
5	150	100	5	42,65
6	150	100	15	52,21
7	150	300	5	51,01
8	150	300	15	50,55
9	100	200	10	31,06
10	100	200	10	31,32
11	100	200	10	30,12

$$Y = 30,833 + 2,225x_1 + 2,485x_2 + 1,923x_3 - 1,367x_1x_2 - 0,405x_1x_3 - 2,045x_2x_3 + 0,344x_3^2 \dots\dots\dots (1)$$

and

$$x_i = \frac{X_i - X_i^*}{\Delta X_i} \dots\dots\dots (2)$$

where x_i is the coded value of the ith independent variable, X_i is the uncoded value of the ith independent variable at the centre point and Δ X_i is the step change value.

The fit of the model was expressed by the determination of coefficient R², which was calculated to be 0.991, this indicated that 1.0% of total variation not explained by the model. The observed values Y_o and predicted values Y_p is given in Table 3. The optimal values of tested variables in coded unit were found to be x₁ = 0.136 ; x₂ = 1.293 and x₃ = 1.128 corresponding to the optimal concentration of total sugar, concentration of nitrogen and inoculum size of 106.8 g/L, 329.3 mg N/L and 15.64% v/v, respectively.

Table 3. Observed responses and predicted values

Run	Sugar	Nitrogen	Inoculum	Yield (%)	
	g/L	mg N/L	% v/v	Y _o	Y _p
1	50	100	5	21.02	20.73
2	50	100	15	29.18	29.47
3	50	300	5	32.23	32.52
4	50	300	15	33.38	33.09
5	150	100	5	28.43	28.72
6	150	100	15	34.80	35.05
7	150	300	5	34.40	34.63
8	150	300	15	33.70	33.99
9	100	200	10	31.06	30.83
10	100	200	10	31.32	30.83
11	100	200	10	30.12	30.83

Y_o = observed yield ; Y_p = observed yield

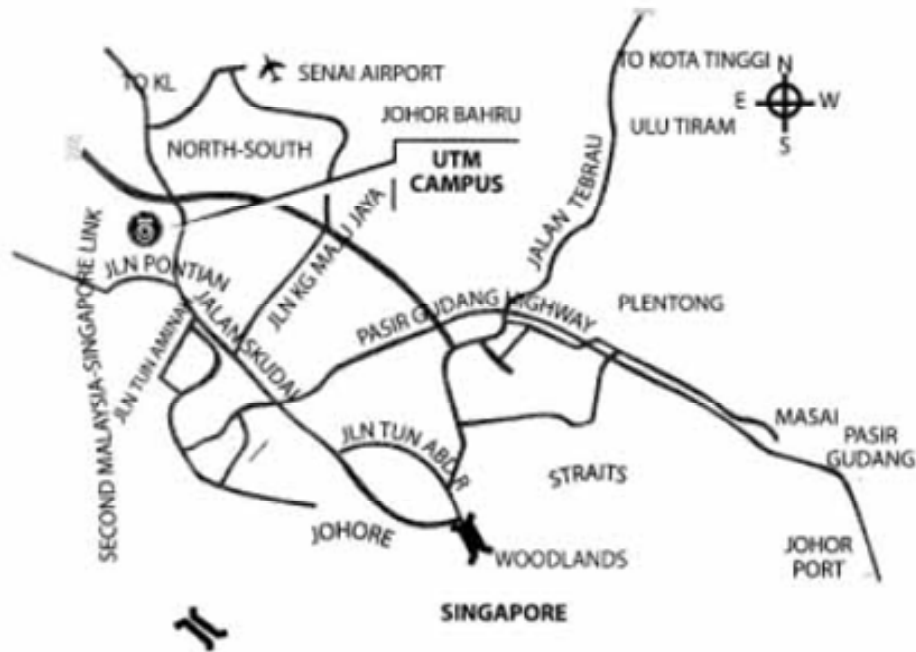
4. Conclusions

The pineapple waste juice can be used as carbon source for ethanol production by *Saccharomyces cerevisiae*. The optimum conditions for pineapple waste juice fermentation were concentration of total sugar, 106.8 g/L; concentration of nitrogen, 329.3 mg N/L; and inoculum size, 15.64% v/v.

References

- [1] Badan Pusat Statistik Indonesia. 2000. Production of Fruits, *Horticulture statistics*.
- [2] Hadi, P.U. 2001. The Case Study of Canned Pineapple in Indonesia, *Center of Agro-socio Economist Research and Development*, Ministry of Agriculture.
- [3] Buckle, K.A. 1989. Biotechnology Opportunities in Waste Treatment and Utilisation for the Food Industry. In *Biotechnology and The Food Industry*, Ed. Rogers P.L. Breach Science Publishers. New York.
- [4] Bardiya, N.; Somayaji, D.; and Khanna, N. 1996. Biomethanation of Banana Peel and Pineapple Waste. *Bioresource Technology*, 58.
- [5] Sasaki, K., Noparatnaraphorn, N. And Nagai, S. 1991. Use of Phytosynthetic Bacteria for the Production of Single Cell Protein and Chemicals from Agro Industrial Waste., In *Bioconversion of Waste Material to Industrial Product*. Ed. Martin, A.M. Elviser Applied Science, London .
- [6] Vimal, O.P.; and Adsole, P.G. Utilisation of Fruit and Vegetables Waste. *Research and Industry* 21(1).
- [7] Abdullah, M.B. Haryani K. Dan Ashila R.M. 2003. Studi Awal Pembuatan Asam Sitrat dari limbah nanas Menggunakan *A. Niger* . *Proseeding Seminar Rekayasa Kimia dan Proses* . Universitas Diponegoro. Semarang. Indonesia.
- [8] Abdullah, M.B. 2002. Lactic Acids Fermentation of Pineapple Waste Using *L. delbrueckii*. Ph.D *Dissertation*. Department of Chemical Engineering . Universiti Teknologi Malaysia.
- [9] Lazaro, M.J. 1989. Liquid Chromatography Determination of Acid and Sugar in Homolactic Cucumbar Fermentation. *Journal AOAC* 72(1).
- [10] Haaland, P.D. 1989. *Experimental Design in Biotechnology*. New York, Marcel Dekker.Inc.

LOCATION





The Organizing Committee would like to express our heartfelt thanks and greatest appreciation

To

***Deputy Vice Chancellor (Academic) of Universiti Teknologi Malaysia
Prof Ir Dr Siti Hamisah Binti Tapsir***

for graciously officiating the opening ceremony of the
1st Regional Postgraduate Conference on Engineering and Science 2006

&

***Prof. Dr. Rahmawan Ahamad
Dean, School of Graduate Studies, Universiti Teknologi Malaysia***

For Officiating The Closing Ceremony.

We would also like to take this opportunity to express our greatest appreciation and thanks to the following:

***Staff and Students, School of Graduate Studies, UTM
Faculty of Education, UTM
Indonesian Student Association
All Speakers and Paper Presenters
Chairpersons and Rapporteurs
Sponsors***

And everyone who has given their time and invaluable assistance in organizing this auspicious and successful event.



