Journal Information

Maejo International Journal of Science and Technology (ISSN 1905-7873 © 2011), the international journal for preliminary communications in Science and Technology is the first peerrefereed scientific journal of Maejo University (www.mju.ac.th). Intended as a medium for communication, discussion, and rapid dissemination of important issues in Science and Technology, articles are initially published online in an open access format, which thereby gives authors the chance to communicate with a wide range of readers in an international community.

Publication Information

MIJST is published triannually. Articles are available online and can be accessed free of charge at http://www.mijst.mju.ac.th. Printed and bound copies of each volume are produced and distributed to authors and selected groups or individuals. This journal and the individual contributions contained in it are protected under the copyright by Maejo University.

Abstracting/Indexing Information

MIJST is covered and cited by Science Citation Index Expanded, SCOPUS, Journal Citation Reports/Science Edition, Zoological Record, Chemical Abstracts Service (CAS), SciFinder Scholar, Directory of Open Access Journals (DOAJ), CAB Abstracts, ProQuest, and Google Scholar.

Contact Information

Editorial office: Maejo International Journal of Science and Technology (MIJST), 1st floor, Orchid Building, Maejo University, San Sai, Chiang Mai 50290, Thailand

Tel: +66-53-87-3880
E-mails: duang@mju.ac.th

MAEJO INTERNATIONAL JOURNAL OF SCIENCE AND TECHNOLOGY

MAEJO INTERNATIONAL JOURNAL OF SCIENCE AND TECHNOLOGY

The International Journal for the Publication of Preliminary Communications in Science and Technology

http://www.mijst.mju.ac.th
ISSN 1905-7873

MAEJO INTERNATIONAL JOURNAL
OF SCIENCE AND TECHNOLOGY

Volume 6, Issue 2 (May - August 2012)

CONTENTS

http://www.mijst.mju.ac.th
ISSN 1905-7873

MAEJO INTERNATIONAL JOURNAL OF SCIENCE AND TECHNOLOGY

Volume 6, Issue 2 (May - August 2012)

Author Index

Instructions for Authors

A proper introductory e-mail page containing the title of the submitted article and certifying its originality should be sent to the editor (Duang Buddhasukh, e-mail : duang@mju.ac.th). The manuscript proper together with a list of suggested referees should be attached in separate files. The list should contain at least 3 referees (one or more non-native) with appropriate expertise. Each referee's academic/professional position, scientific expertise, affiliation and e-mail address must be given. The referees should not be affiliated to the same university/institution as any of the authors, nor should any two referees come from the same university/institution.

Failure to conform to the above instructions will result in non-consideration of the submission.

Please also ensure that English and style is properly edited before submission. UK style of spelling should be used. Authors who would like to consult a professional service can visit www.journalexperts.com, www.proof-reading-service.com, www.editage.com, www.enago.com, www.bioedit.co.uk (bioscience and medical papers), www.bioscienceeditingsolutions.com, www.scribendi.com, www.letpub.com, or www.papersconsulting.com.

Important : Manuscript with substandard English and style will not be considered.

Warning : Plagiarism (including self-plagiarism) may be checked for at *the last* stage of processing and, if detected, will result in a rejection and blacklisting.

Manuscript Preparation

Manuscripts must be prepared in English using a word processor. MS Word for Macintosh or Windows, and .doc or .rtf files are preferred. Manuscripts may be prepared with other software provided that the full document (with figures, schemes and tables inserted into the text) is exported to a MS Word format for submission. Times or Times New Roman font is preferred. The font size should be 12 pt and the line spacing 'at least 17 pt'. A4 paper size is used and margins must be 1.5 cm on top, 2.0 cm at the bottom and 2.0 cm on both left and right sides of the paper. Although our final output is in .pdf format, authors are asked NOT to send manuscripts in this format as editing them is much more complicated. Under the above settings, a manuscript submitted should not be longer than **15 pages** for a full paper or **20 pages** for a review paper.

A template file may be downloaded from the *Maejo Int. J. Sci. Technol.* homepage. (DOWNLOAD HERE)

Authors' full mailing addresses, homepage addresses, phone and fax numbers, and e-mail addresses homepages can be included in the title page and these will be published in the manuscripts and the Table of Contents. The corresponding author should be clearly identified. It is the corresponding author's responsibility to ensure that all co-authors are aware of and approve of the contents of a submitted manuscript.

A brief (200 word maximum) Abstract should be provided. The use in the Abstract of numbers to identify compounds should be avoided unless these compounds are also identified by names.

A list of three to five keywords must be given and placed after the Abstract. Keywords may be single words or very short phrases.

Although variations in accord with contents of a manuscript are permissible, in general all papers should have the following sections: Introduction, Materials and Methods, Results and Discussion, Conclusions, Acknowledgments (if applicable) and References.

Authors are encouraged to prepare Figures and Schemes in colour. Full colour graphics will be published free of charge.

Tables and Figures should be inserted into the main text, and numbers and titles supplied for all Tables and Figures. All table columns should have an explanatory heading. To facilitate layout of large tables, smaller fonts may be used, but in no case should these be less than 10 pt in size. Authors should use the Table option of MS Word to create tables, rather than tabs, as tab-delimited columns are often difficult to format in .pdf for final output. Figures, tables and schemes should also be placed in numerical order in the appropriate place within the main text. Numbers, titles and legends should be provided for all tables, schemes and figures. Chemical structures and reaction schemes should be drawn using an appropriate software package designed for this purpose. As a guideline, these should be drawn to a scale such that all the details and text are clearly legible when placed in the manuscript (i.e. text should be no smaller than 8-9 pt).

For bibliographic citations, the reference numbers should be placed in square brackets, i.e. [ ], and placed before the punctuation, for example [4] or [1-3], and all the references should be listed separately and as the last section at the end of the manuscript.

Format for References

Journal :

1. D. Buddhasukh, J. R. Cannon, B. W. Metcalf and A. J. Power, ″Synthesis of 5-n-alkylresorcinol dimethyl ethers and related compounds *via* substituted thiophens", *Aust. J. Chem.*, **1971**, *24*, 2655-2664.

Text :

2. A. I. Vogel, "A Textbook of Practical Organic Chemistry", 3rd Edn., Longmans, London, **1956,** pp. 130-132.

Chapter in an edited text :

3. W. Leistritz, "Methods of bacterial reduction in spices ", in "Spices: Flavor Chemistry and Antioxidant Porperties" (Ed. S. J. Risch and C-T. Ito), American Chemical Society, Washington, DC, **1997**, Ch. 2.

Thesis / Dissertation :

4. W. phutdhawong, “Isolation of glycosides by electrolytic decolourisation and synthesis of pentinomycin”, *PhD. Thesis*, **2002**, Chiang Mai University, Thailand.

Patent :

5. K. Miwa, S. Maeda and Y. Murata, "Purification of stevioside by electrolysis", *Jpn. Kokai Tokkyo Koho 79 89,066* (**1979**).

Proceedings :

6. P. M. Sears, J. Peele, M. Lassauzet and P. Blackburn, "Use of antimicrobial proteins in the treatment of bovine mastitis", Proceedings of the 3rd International Mastitis Seminars, **1995**, Tel-Aviv, Israel, pp. 17-18.

<u>Manuscript Revision Time</u>

Authors who are instructed to revise their manuscript should do so within **45** days. Otherwise the revised manuscript will be regarded as a new submission.

Maejo Int. J. Sci. Technol. **2012**, *6*(02), 159-185
Maejo International
Journal of Science and Technology

ISSN 1905-7873
Available online at www.mijst.mju.ac.th

Full Paper

Modelling and analysis of the effect of stacking chips with TSVs in 3D IC package encapsulation process

Chu Yee Khor* and Mohd Zulkifly Abdullah

School of Mechanical Engineering, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Penang, Malaysia

* Corresponding author, e-mail: cykhor_1985@hotmail.com

Received: 6 April 2011 / Accepted: 29 April 2012 / Published: 1 May 2012

Abstract: This paper presents the modelling and analysis of the encapsulation process for three-dimensional (3D) stacking-chip package with through-silicon via (TSV) integration. The fluid-structure interaction of the 3D stacking-chip package encapsulation was modelled by finite volume and finite element codes, which were solved separately. The effect of the increase in the number of stacking chips was analysed. The visualisation of the 3D stacking-chip package encapsulation process was presented at different filling times. The void formation around the stacking chips was identified for each case. The displacement and von Mises stress for the copper through-silicon vias were determined. The use of designed inlet-outlet heights in the integrated circuit package maintained the filling time of the encapsulation process and reduced the void of the packages as the number of stacking chips increased. The encapsulation model facilitated a clear visualisation and enhanced fundamental understanding of the design of a 3D integrated circuit encapsulation. The proposed analysis is expected to be a reference and guide in the design and improvement of 3D integration packages.

Keywords: 3D IC integration, through-silicon vias (TSVs), stacking chips, fluid-structure interaction, finite volume, finite element

---

INTRODUCTION

In the microelectronic industry, the trend of integrated circuit (IC) packaging is now towards miniaturisation and establishment of high-capacity and high-performance circuits for the rapid development of electronic devices. Three-dimensional (3D) integration [1] with through-silicon vias (TSVs) and 3D IC packaging [2] technologies enable package designers to accomplish their goals through stacking chips with TSVs. In 3D IC integration, thinned silicon chips and micro-size interconnection require reliable housings to maintain the device's reliability and protect them from the hazardous environment. The encapsulation process [3, 4] using epoxy-moulding compound (EMC) was widely used for various IC packages such as thin quad flat packages (TQFPs) [5, 6], thin profile small outline packages with wide side of lead on chip (TSOP II 54 L LOC) [7, 8], stacking-chip scale packages [9], mould array packages [10] and moulded underfills [11]. During the encapsulation process, the EMC is fed into the cavity through transfer moulding to encapsulate structures such as the silicon chips, wire bondings, solder bumps, paddles and passive components. An improper encapsulation process could lead to the reduction in package reliability and defects such as warpage, interconnector fracture, critical displacement or deformation of chips, solder bumps, TSVs and void formation.

Publications regarding TSVs report on studies of thermal stress, thermomechanical reliability, TSV filling, thinned stacking wafers, wafer-level bonding, TSV structures, strength evaluation, electro-migration performance, and stacking memory chips with TSVs. Lau [1] reported a considerable number of studies on TSV manufacturing and its hidden costs, focusing on TSVs for 3D integration. Equations for manufacturing cost per good die and TSV calculation were also discussed. Lau also addressed important recommendations and development of TSVs. Tu [2] reported the reliability challenges in the 3D IC packaging technology. Joule heating was a major concern, and electro-migration, stress migration and chip warpage under compressive stress were also considered challenging issues. Ko and Chen [12] conducted an excellent review on wafer-level bonding for 3D integration including advantages and disadvantages of the bonding technology. Wafer surface contamination, which was potentially due to the fabrication tool, was also a primary concern. High via density and better alignment could be achieved by silicon direct bonding technology and mechanical stability could be achieved by hybrid bonding.

An advantage of TSV technology is its application in diverse packages. Falat et al. [13] applied TSVs in microelectromechanical systems (MEMS) through the direct face-to-face chip-to-wafer bonding of an MEMS device and focused on the mechanical aspects of wafer-level packaging. They employed the finite-element method (FEM) on TSVs in substrates, considering temperature excursion, pressure during moulding, materials, and handling load effect on mechanical stress. Wang et al. [14] also proposed and developed an integrated MEMS probe card for wafer-level IC testing. The TSV composite structure utilised a silicon cantilever, guaranteeing better signal routing and low path resistance. In their study, no degradation was observed in the 100,000-cycle lifetime test.

The TSV technology is also applicable in complementary metal oxide semiconductors (CMOS). Gagnard and Mourier [15] reported that the fabrication of CMOS imager sensor has been directed towards 3D integration through bonding and via filling methods. The challenge in 3D integration is the mounting of the double-sided device on the interposers, which allows the connection of chips at various locations and sizes. Abouelatta-Ebrahim et al. [16] implemented the finite difference method in 2D and 3D TSV simulations and investigated the sensitive layer of CMOS inverters and stacking devices in 3D circuits.

The IC packages became more compact and had higher performance and capacity as TSVs developed. Thinned wafers or silicon dies can now be stacked and their sizes reduced, as compared with those in conventional packaging. Ohba et al. [17] developed thinned wafer stacking and copper multilevel interconnects. They found that the increase in the number of stacking wafers was dependent on the degree of integration. The method developed was expected to be a low-cost and high-integration technology for post scaling. Burghartz et al. [18] reviewed the ultra-thin chip technology and its applications and presented a new challenge: the performance and integration density limitation in flexible electronics. Simulation modelling played a significant role in the TSV investigations. The FEM and modified virtual crack closure technique were utilised by Wu et al. [19] to investigate the delaminating behaviour in 3D stacking-chip package (SCP), focusing on the stacking copper bumps. They reported that the mesh density only affected the G value (energy release rate) at a specific element size.

Diverse TSV applications require different designs. Thus, various parameters should be considered and optimised in the TSV design process. However, the effects of TSV design parameters could reduce package reliability. Landani [20] analysed the effect of design parameters on the thermo-mechanical properties of TSVs and solder joints in 3D ICs and discovered that the underfill thickness and stiffness significantly affects the durability of the solder joint and that the diameter is an important factor for TSV durability. An increase in die thickness significantly decreased the durability of the solder joint and TSVs. Moreover, Landani [21] studied the effects of design parameters on the solid-liquid interdiffusion using FEM and considered the elastic-plastic behaviour in the finite element analysis. The most influential factors for interface stress and copper interconnectors were the silicon die and substrate thickness. Meanwhile, die thickness and underfill stiffness were identified as factors influencing the stress of solid-liquid interdiffusion bonds. The effect of the interconnection size and pitch on stress was smaller than that of die thickness.

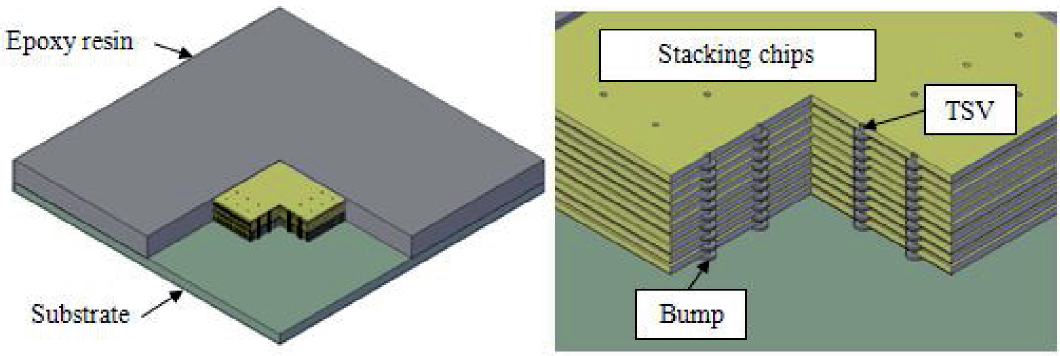
The conductor resistance generally varies with the diameter. Therefore, this aspect should also be considered in TSV research. Kamto et al. [22] investigated resistance in single via and long via chains and reported that TSVs are more resistant to temperature cycling and less resistant to higher temperatures than to wire bonding. A different TSV design, advanced clamped-TSV, was developed by Shen et al. [23] and was used in 3D chip-to-chip/wafer packaging. This cost-effective TSV fabrication process does not have layer deposition and photo processes. The clamped-TSV yielded lower thermal stress and higher bonding reliability for stacked bonding compared with the typical TSV design. The electromigration performance of TSVs was studied by Tan et al. [24] using FEM on a different pattern of top metallisation. According to them, the thermomechanical stress, which behaves asymmetrically, is more dominant over electromigration than the current density and failure modes of electromigration in TSVs.

Package reliability has always been a concern among researchers in the microelectronic industry. To achieve their goals, these researchers consider many aspects in the research and development of TSV technology. Wu et al. [25] employed FEM and a four-point bending test to evaluate the strength of silicon dies covered with polymer films. They found that polymer films enhanced the strength of 10 layers of 3D stacking dies during the fabrication process; the structure had no die cracks during thermal cycling. Tsai and Chen [26] proposed a new point load test on the different die surfaces to evaluate the strength of silicon dies. The point load test provided more data than the four- and three-point tests when considering the effect of surface roughness on silicon die.

The various IC applications of TSV technology and the diverse applications of the encapsulation process in different packages were reported [3-11]. However, the investigation of the encapsulation process in 3D IC packages considering TSVs and structural analysis is still rarely reported in the literature. The encapsulation process of stacking-die packages reported by Moon et al. [4] and Abdullah et al. [9] mainly focused on the fluid flow analysis and the interconnection of their IC packages using wire bonding instead of TSVs. Hence, the present study considers the 3D IC package with stacking chips connected by TSVs. It also focuses on the 3D modelling of the SCP and the effect of inlet-outlet heights during the encapsulation process. FLUENT 6.3 and ABAQUS 6.9 are utilised to perform the fluid and structural analyses. The effects of the EMC flow behaviour in the stacking chips and the stress and displacement on the TSVs are also investigated.

PROBLEM DESCRIPTION

The major issues in IC packaging are the IC package malfunction due to cracks and fractures at the interconnector, the critical displacement of the silicon chips and interconnector, and the void formation. Factors that influence the package quality during the encapsulation process include package design, material selection and process control. The design of the physical characteristics of the IC package and moulding parameters, such as inlet and outlet gates, orientation of stacking chips, number of stacking chips, solder bump standoff height and package size, may affect the fluid flow, filling time and void formation in the encapsulation [6]. Improper selection of EMC materials [11] may also cause void formation, delamination and shrinkage of compound. Improper process control such as that of moulding pressure and temperature [27] may also affect the package reliability. Figure 1 illustrates a schematic diagram of 10 layers of the SCP with TSVs. Generally, the increase in the number of stacking chips also increases the package volume. Therefore, a longer filling time is required and void formation may occur in the package, especially at the narrow space between stacking chips. Thus, the moulding parameters should be properly designed to maintain package reliability and reduce manufacturing cost. The present study mainly focuses on the effect of the number of stacking chips (2-6 stacking chips) on the fluid flow and structural deformation of the interconnector. Other parameter issues are considered as beyond the scope of the study. The designed inlet ($h_{inlet} = 0.5t$) and outlet ($h_{outlet} = 0.25t$) heights are applied in the package design, where t is the thickness of the package. The current study investigates five cases, namely cases 1–5 for packages with 2, 3, 4, 5 and 6 stacking chips respectively. The EMC flow behaviour is modelled using the Castro-Macosko model which was programmed using Microsoft VISUAL STUDIO.NET [6] and compiled with FLUENT code using user-defined functions. The force induced by the viscous fluid on the silicon chips and TSVs is used in the structural analysis in the ABAQUS code.



(a) Isometric view (b) Detailed view

Figure 1. Isometric and detailed views of SCP

METHODS

Nomenclature

A_1, A_2	Pre-exponential factors	1/s
B	Exponential-fitted constant	Pa.s
C_1,C_2	Fitting constant	-
C_P	Specific heat	J/kg-K
E_1,E_2	Activation energies	K
E	Elastic modulus	GPa
F	Front advancement parameter	-
$\vec{g}$	Gravitational acceleration	m/s^2
k	Thermal conductivity	W/m-K
K_1, K_2	Rate parameters described by an Arrhenius temperature dependency	1/s
m_1, m_2	Constants for the reaction order	-
n	Power law index	-
p	Pressure	Pa
T	Temperature	K
t	Time	s
T_b	Temperature-fitted constant	K
u	Fluid velocity component in x-direction	mm/s
v	Fluid velocity component in y-direction	mm/s
w	Fluid velocity component in z-direction	mm/s
x, y, z	Cartesian coordinates	-
Greek letters		
α	Conversion of reaction	-
α_{gel}	Degree of cure at gel	-
ΔH	Exothermic heat of polymerisation	J//kg
η	Viscosity	Pa.s
η_0	Zero shear rate viscosity	Pa.s

ρ	Density	kg/m^3
ρ_s	Solid density	kg/m^3
$\overleftrightarrow{\sigma}$	Recoverable stresses	Pa
τ	Shear stress	Pa
$\dot{\gamma}$	Shear rate	1/s
τ^*	Parameter that describes the transition region between zero shear rates and the power law region of the viscosity curve	Pa
ν	Poisson ratio	-

Governing Equations

In fluid flow analysis the continuity, Navier-Stokes, and energy equations are accounted for in describing the EMC flow by considering the fluid incompressible. The equations are solved in the simulation (FLUENT) as follows.

Continuity equation:

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0 \tag{1}$$

The motion of the fluid flow in the encapsulation process is described by Navier-Stokes equation [6], expressed in terms of x, y and z directions:

x-direction

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}+w\frac{\partial u}{\partial z}=-\frac{1}{\rho}\frac{\partial p}{\partial x}+\left[\frac{\partial}{\partial x}\left(\eta\frac{\partial u}{\partial x}\right)+\frac{\partial}{\partial y}\left(\eta\frac{\partial u}{\partial y}\right)+\frac{\partial}{\partial z}\left(\eta\frac{\partial u}{\partial z}\right)\right] \tag{2}$$

y-direction

$$\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}+w\frac{\partial v}{\partial z}=-\frac{1}{\rho}\frac{\partial p}{\partial y}+\left[\frac{\partial}{\partial x}\left(\eta\frac{\partial v}{\partial x}\right)+\frac{\partial}{\partial y}\left(\eta\frac{\partial v}{\partial y}\right)+\frac{\partial}{\partial z}\left(\eta\frac{\partial v}{\partial z}\right)\right] \tag{3}$$

z-direction

$$\frac{\partial w}{\partial t}+u\frac{\partial w}{\partial x}+v\frac{\partial w}{\partial y}+w\frac{\partial w}{\partial z}=-\frac{1}{\rho}\frac{\partial p}{\partial z}+\left[\frac{\partial}{\partial x}\left(\eta\frac{\partial w}{\partial x}\right)+\frac{\partial}{\partial y}\left(\eta\frac{\partial w}{\partial y}\right)+\frac{\partial}{\partial z}\left(\eta\frac{\partial w}{\partial z}\right)\right] \tag{4}$$

The temperature of fluid flow in the encapsulation process is described using the energy equation [6]:

$$\rho C_p\left(\frac{\partial T}{\partial t}+u\cdot\nabla T\right)=\nabla\left(k\nabla T\right)+\Phi, \tag{5}$$

where the source term Φ is defined as:

$$\Phi = \eta \dot{\gamma}^2 \tag{6}$$

$$\dot{\gamma} = \sqrt{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 + \left(\frac{\partial w}{\partial z}\right)^2} \tag{7}$$

The viscosity of the EMC during encapsulation is modelled by the Castro-Macosko model [5, 28-30] with curing effect (Kamal's equation) in the current study. The viscosity changes and reactions of the encapsulant are influenced by the temperature in the process. The encapsulant material is assumed to be a general Newtonian fluid. The Castro-Macosko model is used to predict the relationship between viscosity and degree of polymerisation and describes the viscosity of the encapsulant material:

$$\eta(T, \dot{\gamma}) = \frac{\eta_0(T)}{1 + \left(\frac{\eta_0 \dot{\gamma}}{\tau^*}\right)^{1-n}} \left(\frac{\alpha_g}{\alpha_g - \alpha}\right)^{C_1 + C_2 \alpha} \tag{8}$$

with

$$\eta_0(T) = B \exp\left(\frac{T_b}{T}\right) \tag{9}$$

where B is an exponentially fitted constant, T is temperature, T_b is a temperature-fitted constant, n is the power law index, η_o is the zero-shear viscosity, $\dot{\gamma}$ is the shear rate, τ^* describes the transition region between zero-shear rate and power law region of the viscosity curve, α is the conversion of reaction, α_g is degree of cure at gel, and C_1 and C_2 are fitting constants.

Kamal's equation [5, 28-30] is coupled with the Castro-Macosko model in the current study. The chemical conversion of EMC occurs during the IC encapsulation process. The cross-linking reactions of the EMC material increase its viscosity and influence the conversion of the reaction (α). The consideration of the curing effect, which is described by Kamal's equations, is important in describing the EMC curing behaviour in IC encapsulation process. The rate of chemical conversion of the EMC in this model is predicted as follows:

$$\frac{d\alpha}{dt} = \left(k_1 + k_2 \alpha^{m_1}\right)\left(1 - \alpha\right)^{m_2} \tag{10}$$

$$k_1 = A_1 \exp\left(-\frac{E_1}{T}\right) \tag{11}$$

$$k_2 = A_2 \exp\left(-\frac{E_2}{T}\right) \tag{12}$$

where α is the degree of conversion, A_1 and A_2 are Arrhenius pre-exponential factors, E_1 and E_2 are activation energies, m_1 and m_2 are reaction orders, T is the absolute temperature, and k_1 and k_2 are rate parameters described by Arrhenius temperature dependency.

The volume of fluid (VOF) model [6] is applied to treat the two distinct phases (air and EMC). It is used to calculate the interface tracking of the encapsulant/air interface in the analysis. The computational domain consists of both the encapsulant and air regions. As time increases, the encapsulant phase increases and the air phase decreases. The melt front over time is governed by the following transport equation:

$$\frac{dF}{dt} = \frac{\partial F}{\partial t} + u\frac{\partial F}{\partial t} + v\frac{\partial F}{\partial t} + w\frac{\partial F}{\partial t} - \left\{ \frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + \frac{\partial^2 F}{\partial z^2} \right\} = 0 \tag{13}$$

where F is the fraction of the cell volume occupied by liquid. In the simulation result of flow front advancement (as shown later in Figure 11), the cell (mesh element) that contains only the resin is represented in red (F=1), and the cells without resin are represented in white (F=0). Values between 0 and 1 ($0<F<1$) in the 'interface' cells are referred as the EMC melt front.

In the structural analysis, the momentum equation [31] used in FEM based (ABAQUS) for solving the structural deformation is as follows:

$$\rho_s \left(\frac{\partial \vec{u}}{\partial t} + \vec{u}.\nabla \vec{u} \right) = -\nabla \overleftrightarrow{\sigma} + \rho_s \vec{b} \tag{14}$$

where ρ_s is density of the solid, $\vec{u}$ is velocity of the solid in *x*-, *y*- and *z*-axes, $\vec{\sigma}$ is recoverable stress, and $\vec{b}$ is mass force.

Simulation Model Description

Generally, the increase in package thickness and volume is attributed to the increase in the number of stacking chips. According to a previous study [6], the rise in the number of inlet gates significantly raises the cavity pressure, reduces the air trap and shortens the filling time in the thin quad flat package (TQFP) encapsulation. However, the present study considers a single inlet and outlet using designed inlet-outlet heights, which are applicable to packages with different numbers of stacking chips. The inlet (h_{inlet}) and outlet (h_{outlet}) heights of the mould were designed (Table 1) to be 50% and 25% of the mould thickness respectively, with a width (w) of 4 mm each. The mould thickness (t) is illustrated in Figure 2, together with the top, front and detailed views of the package with six stacking chips and TSVs. The gap between the mould wall and the top surface of the silicon chips was set to h_1 = 0.1 mm for all cases, as shown in Figure 3. The dimension of the 3D model [19] of the SCP is presented in Table 2.

Table 1. Detail of SCPs with different numbers of stacking chips under study

Case	No. of stacking chips	Mould thickness, t (mm)	h_{inlet} (= 0.5t) (mm)	h_{outlet} (= 0.25t) (mm)
1	2	0.405	0.2025	0.1013
2	3	0.575	0.2875	0.1438
3	4	0.745	0.3725	0.1863
4	5	0.915	0.4575	0.2288
5	6	1.085	0.5425	0.2713

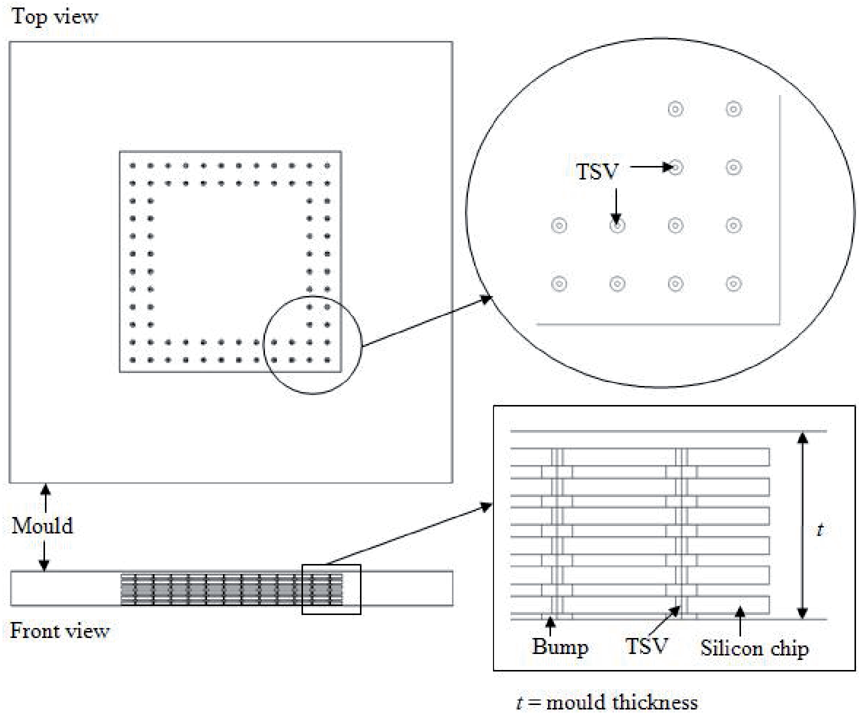


Figure 2. Top view, front view and detailed views of SCP (six chips)

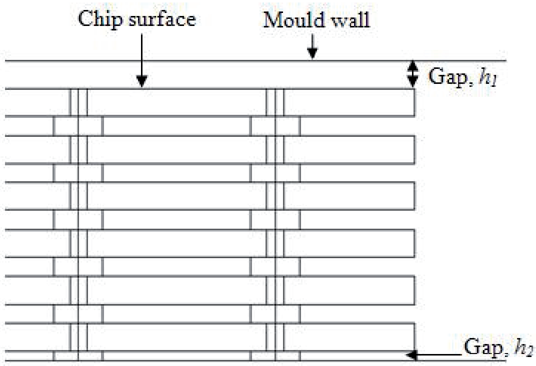


Figure 3. Gaps between mould wall and chip surface

Table 2. Dimension of 3D SCP model [19] used in the FLUENT modelling

Item	Dimension
Silicon chips	10 mm (H) × 10 mm (W) × 0.1 mm(T)
Silicon substrate	20 mm (H) × 20 mm (W) × 0.625 mm (T)
Bump radius	100µm
Bump thickness	35 µm
Via radius	35 µm
Via pitch	800 µm

(H = height; W = width; T = thickness)

FLUENT Modelling

In the analysis, the SCPs were built and meshed according to the dimensions as presented in Table 2. The finite-volume-based software, FLUENT 6.3.26, was utilised to model the fluid flow and solve the governing equation of the EMC flow field in the encapsulation process. A total of 70,000–1,600,000 tetrahedral elements were meshed in the 3D model of the packages, as illustrated in Figure 4.

The epoxy compound, EME 6300HN [5], and air were considered as the phases in the simulation. The Castro-Macosko viscosity model and Kamal's equation were programmed to model the flow behaviour and curing kinetics of the EMC. The VOF technique was applied to track the melt front advancement in the encapsulation, in which the mould temperature (T_w) was set at 170°C with inlet pressure (p_{in}) of 1 MPa and preheat temperature of 90°C. The EMC material properties are summarised in Table 3.

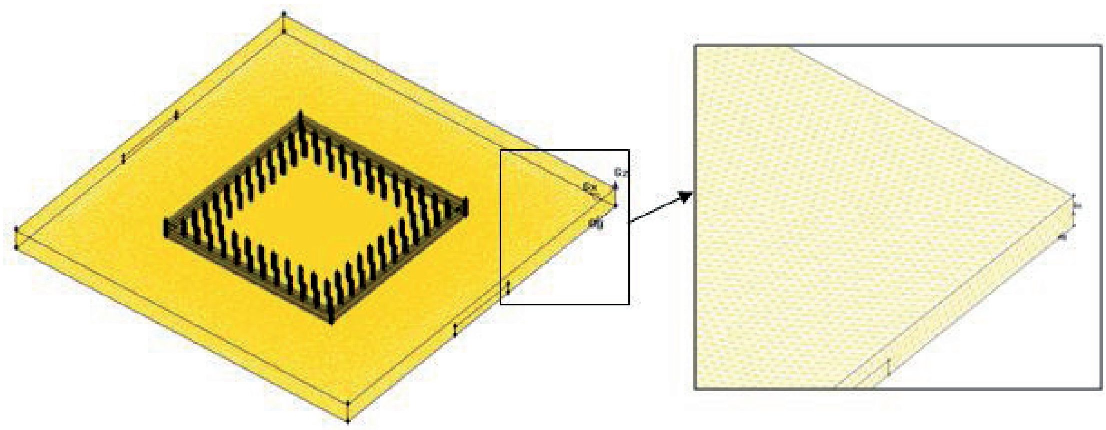


Figure 4. Meshed model for SCP (six chips)

Table 3. EMC material properties [5] used in FLUENT simulation

Castro-Macosko model parameter		Kamal model parameter	
B (kPa.s)	3.81E-7	m_1	1.21
T_b (K)	5230	m_2	1.57
α_g	0.17	A_1 (1/s)	33.53E3
C_1	1.03	A_2 (1/s)	30.54E6
C_2	1.50	E_1 (K)	7161
		E_2 (K)	8589

The first-order upwind discretisation was utilised for both momentum and energy equations, with the SIMPLE scheme for pressure-velocity coupling. A time-dependent formula and an implicit solution were applied for the volume fraction in every time step. The values 'one' and 'zero' were assigned for the volume fractions of the EMC and air phases respectively. However, the user-defined functions were developed using C programming (Microsoft VISUAL STUDIO.NET) and incorporated into the FLUENT simulation to solve the Castro-Macosko model and Kamal's equation. An optimum time step of 0.001 sec. was applied in the simulation and 3400 time steps were taken to fill the mould cavity completely. The simulations were analysed using a computer with Intel Core i3 processor i3-540 at 3.07 GHz and 3.24 GB of RAM.

The boundary conditions of the 3D models were defined in the FLUENT analysis as illustrated in Figure 5. The boundary and initial conditions are as follows:

a) On wall : $u = v = w = 0$; $T = T_w$, $\frac{\partial p}{\partial n} = 0$

b) On centre line : $\frac{\partial u}{\partial z} = \frac{\partial v}{\partial z} = \frac{\partial w}{\partial z} = \frac{\partial T}{\partial z} = 0$

c) On melt front : $p = 0$

d) At inlet : $p = p_{in}(x, y, z)$; $T = T_{in}$

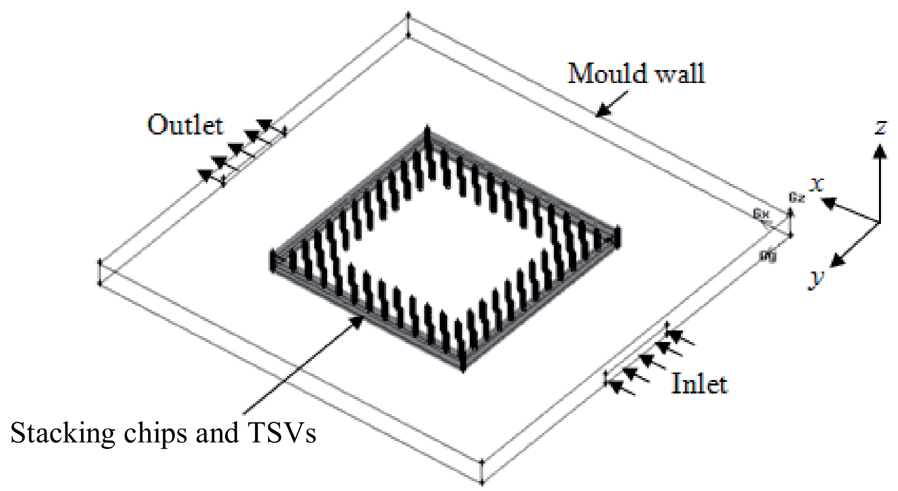


Figure 5. Boundary conditions of the IC package

ABAQUS Modelling

The present structure analysis solves the deformable structure of the TSVs during the encapsulation process using the FEM-based software, ABAQUS. The TSV model was built based on the dimensions in Table 2 and meshed using tetrahedral elements as shown in Figure 6. The structure of the copper TSVs was designated an elastic modulus E = 113.85 GPa and a Poisson ratio ν = 0.35. The mechanical properties of the silicon chip and copper are summarised in Table 4. The fixed boundary condition of the structural model was set at the bottom of the TSV.

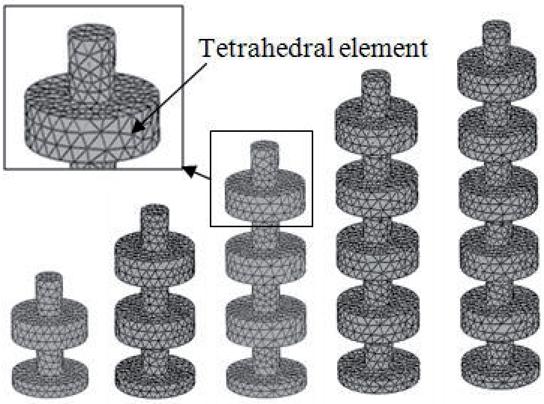


Figure 6. Meshed TSV model for structural analysis using tetrahedral elements

Table 4. Mechanical properties for the FEM analysis [16] using ABAQUS

Material	Elastic modulus, E (GPa)	Poisson ratio, ν
Copper	113.85 [32]	0.35
Silicon	125	0.28

RESULTS AND DISCUSSION

In the present study, the designed inlet-outlet heights were set for the package design as mentioned earlier. The effects of the number of stacking chips on computing time, meshing elements, flow front advancement profile, pressure distribution in the cavity, void formation, drag force, displacement, and stress of the TSVs in the encapsulation process are then discussed.

Model Validation

The predictions of the flow-front advancement for the SCP encapsulation process are impossible because of the lack of experimental data on 3D SCP with TSVs. However, the prediction of the encapsulation and injection mould modelling by the present methodology was well validated in previous studies [6, 33-34]. The prediction of flow-front advancement and viscosity variation matched well with the experimental results. Figure 7 shows a comparison of simulation and experimental results [5] for the TQFP encapsulation process. Both results showed similar filling trends during the process. From the comparison, the strength of FLUENT in handling encapsulation problems was proven. Thus, the predictions of 3D SCP encapsulation could be accomplished using the designed model.

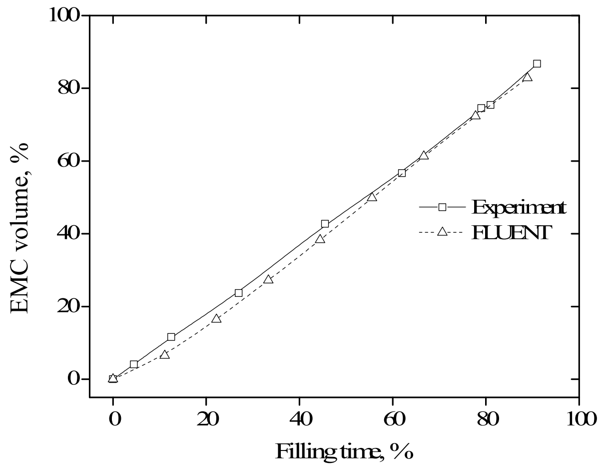


Figure 7. Comparison of simulation and experimental data on EMC volume filled versus filling time during the encapsulation process

Computing Time and Number of Meshing Elements

Generally, the computational fluid dynamics software [6, 35] is widely utilised to handle various engineering problems. The application of this software involves computing time of the simulation, which is always related to the cost in the research work. The modelling and simulation of the complicated 3D model may require more computing time and higher cost. Figure 8 shows that the number of meshing elements is directly proportional to the number of stacking chips in the cases 1–5 (Table 1). As the number of stacking chips increases, the package volume rises, increasing the number of meshing elements in the modelling process. The correlation of computing time and meshing elements is plotted in Figure 9. The computing time is directly proportional to the number of meshing elements in the simulation based on the finite volume method.

Fluid Flow Analysis

The designed inlet-outlet of the IC package was employed for packages with different numbers of stacking chips and increasing inlet and outlet heights. Figure 10 shows the filling volume of the packages versus encapsulation time. Comparison of the filling volume for different cases shows that the increase in number of stacking chips directly increases the package volume. The designed inlet-outlet enabled the volume of the filled packages to follow a similar trend. All packages were almost filled at 3.4 sec. When the inlet gate height was increased, more EMC was fed into the cavity, thus reducing the filling time. The application of the designed inlet-outlet heights (h_{in}=0.5t and h_{out}=0.25t) in the IC package maintained the encapsulation time during the process for the different cases with diverse package volumes and stacking chips at a constant inlet pressure. Furthermore, the design of a multi-inlet gate [6] and optimum inlet location [11] could effectively minimise the filling time in the encapsulation process. Therefore, the process could be controlled through the design of the inlet and outlet areas of the mould cavity.

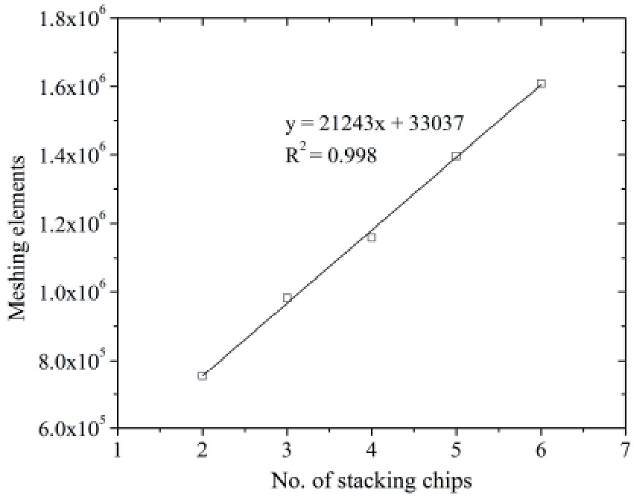


Figure 8. Number of meshing elements in cases 1-5 (No. of stacking chips = 2-6)

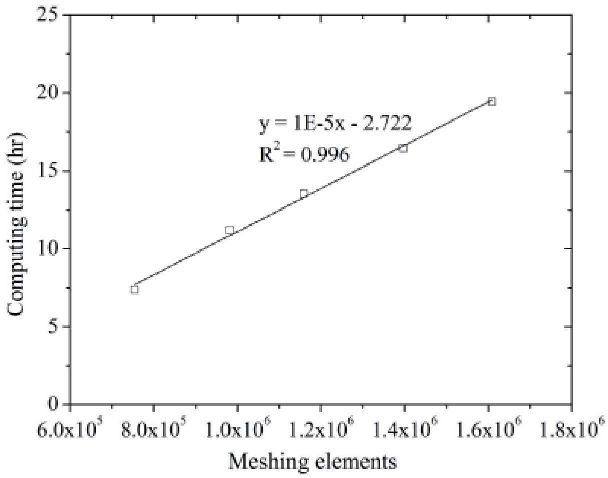


Figure 9. Computing time versus meshing elements

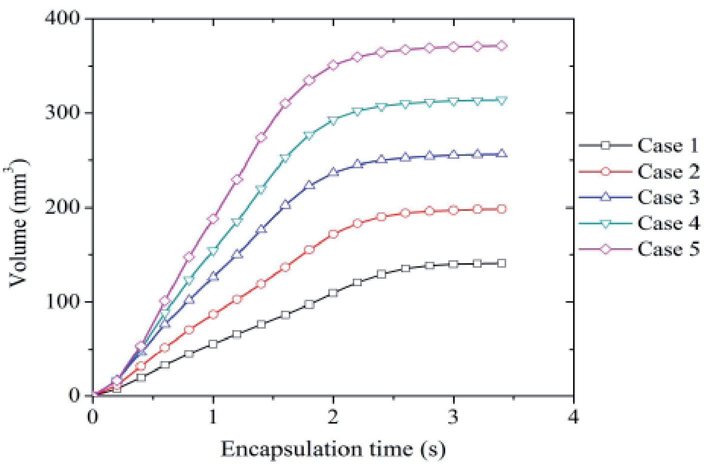
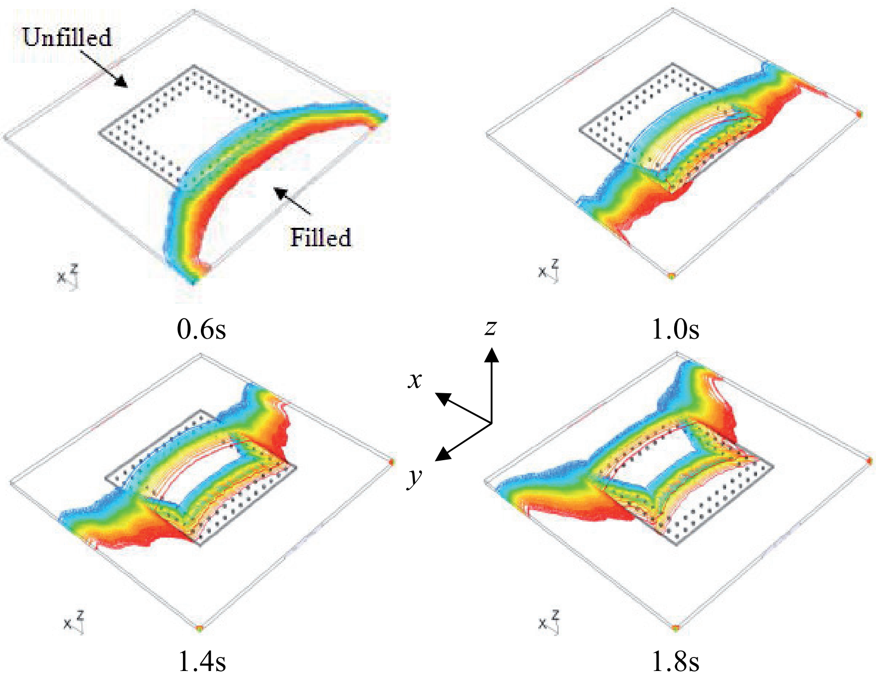


Figure 10. Volume filled during encapsulation

Flow-Front Advancement

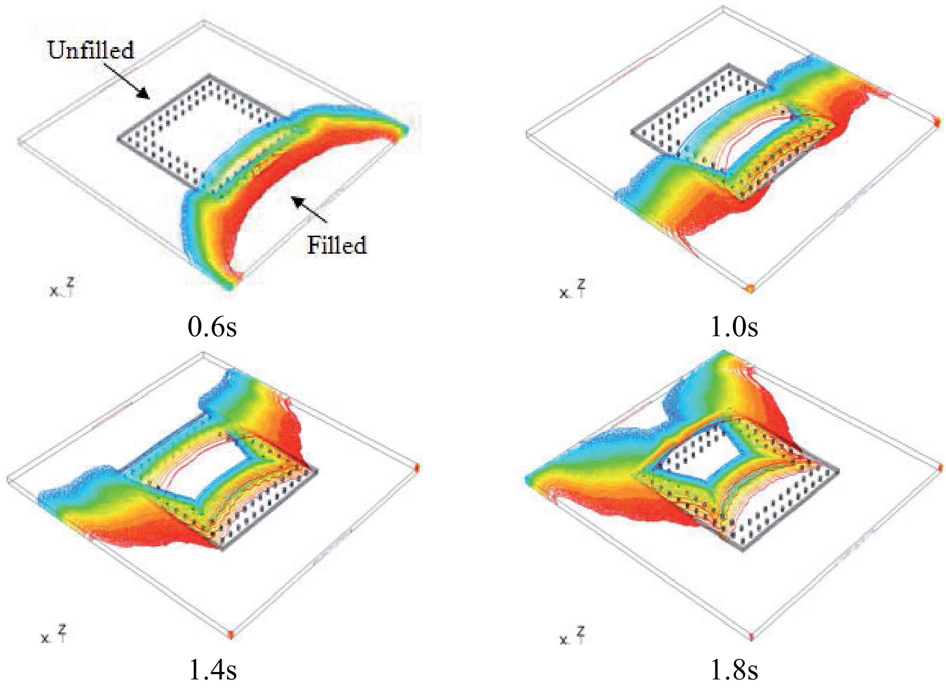
Figure 11 shows the flow-front advancement of the EMC in the SCP at filling times of 0.6, 1, 1.4 and 1.8 sec. during the encapsulation process. As shown, the EMC interacts with the silicon chips and TSVs. The filled and unfilled regions are labelled at 0.6 sec. for each case. The interaction between the moulding compound and the stacking chips occurs at approximately 0.6 sec for each case. Figure 11(a) shows the non-uniform front advancement at 1.0 sec., when the compound flows through gap h_1 (between the top surface of the stacking chips and the mould wall). The flow that passes through gap h_1 is faster than that through the gap between the stacking chips and gap h_2. The small gap between the stacking chips resists the EMC flow and this phenomenon is observed at 1.4 sec. A faster flow at the free regions (no stacking chips) begins at 1.4 sec. until 1.8 sec. The compound completely fills gap h_1 at 1.8 sec. However, unfilled regions still exist in the gap between the stacking chips and gap h_2.

The increase in inlet and outlet areas at a constant inlet pressure significantly influences the flow behaviour (Figures 11(b-e)). More EMC is fed into the cavity with a larger inlet area, as clearly shown at 0.6 sec. The free region in the package increases when the number of stacking chips increases, enabling the EMC to flow more freely and faster than around the stacking chips. This situation is observed in cases 2–5 at filling times of 1.0 and 1.4 sec. A high possibility of void formation around the gap between the stacking chips is identified when the compound totally fills the free regions, as illustrated in Figures 11(b-e) at the filling time of 1.8 sec.

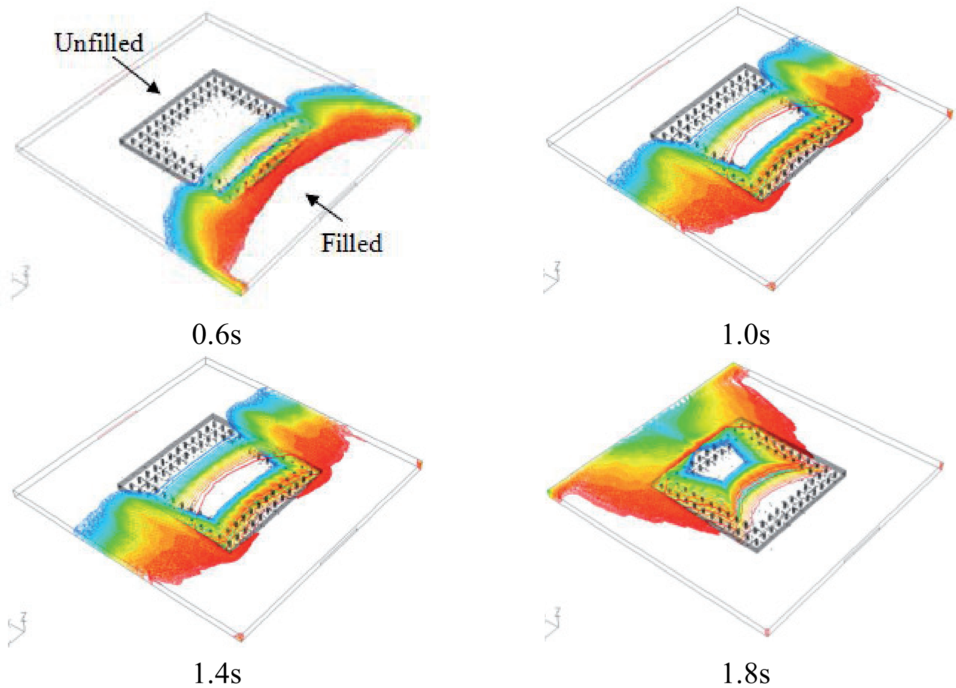


(a) 2 stacking chips (Case 1)

Figure 11. Flow-front advancement of EMC in the packages. Colours in Figure represent volume fractions (0<F<1) of cells (as mentioned in governing equations), which illustrate EMC flow front advancement during the encapsulation process.

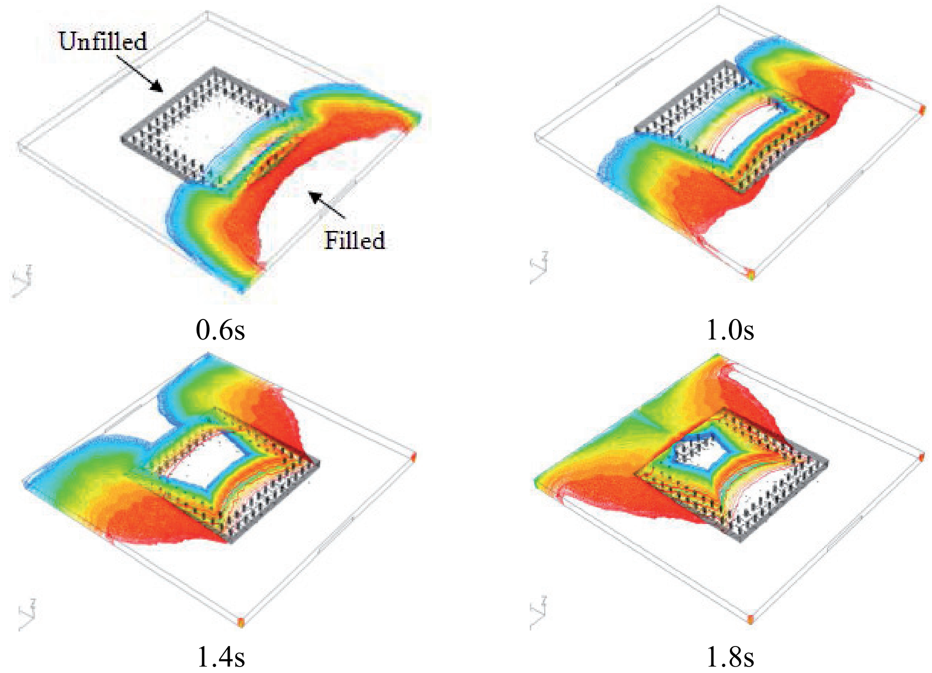


(b) 3 stacking chips (Case 2)

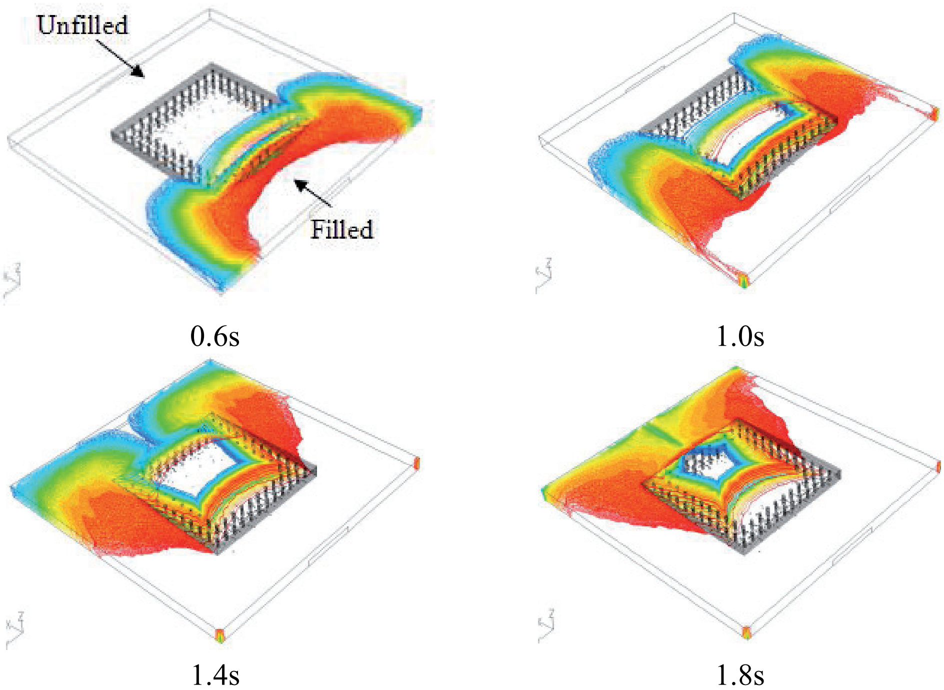


(c) 4 stacking chips (Case 3)

Figure 11 (Cont.)



(d) 5 stacking chips (Case 4)



(e) 6 stacking chips (Case 5)

Figure 11 (Cont.)

Pressure Distribution

The inlet gate design significantly affects the flow front, filled volume and filling time at constant inlet pressure. Therefore, the pressure in the cavity during encapsulation was measured at different positions, as illustrated in Figure 12. The pressure points, P1 to P5, are positioned at the cavity near the stacking chips. P1 and P5 are located at the free regions, whereas P2 and P4 are positioned around the corner of the stacking chips. P3 is positioned between P2 and P4.

Figure 13 shows the measured pressure at P1 to P5 at 2.6 sec. of the encapsulation process before the plunger stops. The pressure variations for all cases show similar trends. The pressure in the cavity is due to the presence of the EMC. As shown in Figure 13, the increase in the inlet height and number of stacking chips raises the pressure in the cavity. This increase may be attributed to the volume of EMC that is fed into the cavity. As the gate height increases, more EMC fills the cavity, thus increasing the cavity pressure. A similar phenomenon was found in the TQFP encapsulation process [6]; the increase in the number of inlet gates raised the cavity pressure. The pressure at P3, which is located in front of the inlet gate, is highest because of a continuous fluid flow from the inlet gate and the obstruction created by the stacking chips. The pressure of the fluid flow at P1 and P5 is lowest because of the absence of stacking chips. The correlations between the inlet gate height (h_{inlet}) and pressure at P1, P2 and P3 are presented in Figure 14. Based on the Figure, the cavity pressure is directly proportional to the inlet gate height.

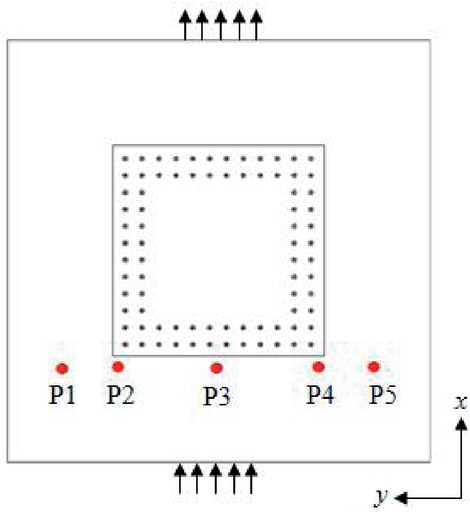


Figure 12. Points P1 to P5 where pressure was measured near the stacking chips

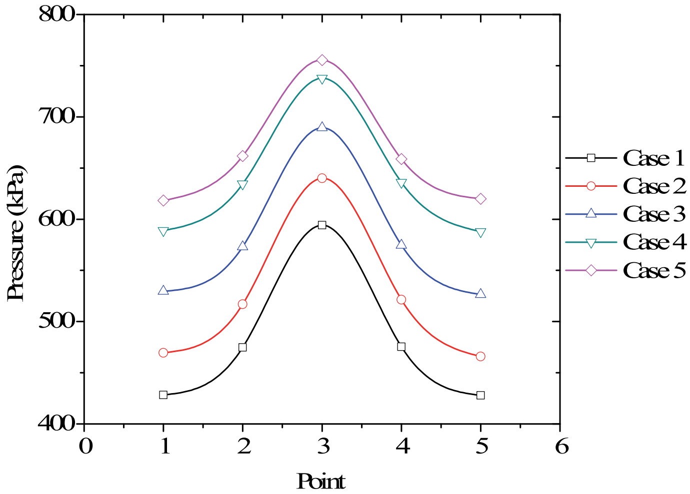


Figure 13. Pressure at points P1–P5 measured at 2.6 sec. of encapsulation for different cases

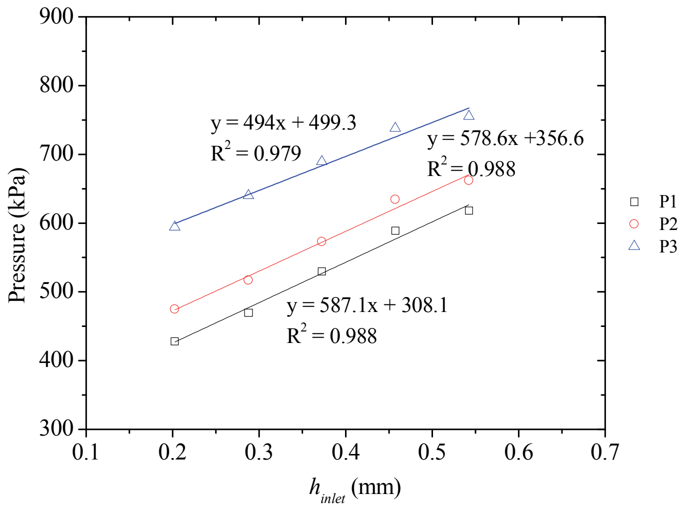


Figure 14. Pressure versus h_{inlet} at P1, P2 and P3

Void Formation

The void formation in the IC package causes the diminution of the package reliability. The number of stacking chips crucially influences the void formation. The voids were formed at the gaps between the stacking chips during encapsulation [31]. To reduce the void in IC packaging, the present study considers the inlet gate height (h_{inlet}). Figure 15 shows the percentage of void formation in the packages for different h_{inlet} values obtained from the current study. As mentioned earlier, h_{inlet} increases as the number of stacking chips increases. The void formation is reduced when h_{inlet} increases as more EMC fills the cavity. The use of a multi-inlet or optimised gate minimises the void formation in the packages [6, 11]. Thus, the use of the designed inlet-outlet heights (h_{inlet} = 0.5t and h_{outlet} = 0.25t) is important for the present stacking-chip IC package design in reducing the void formation.

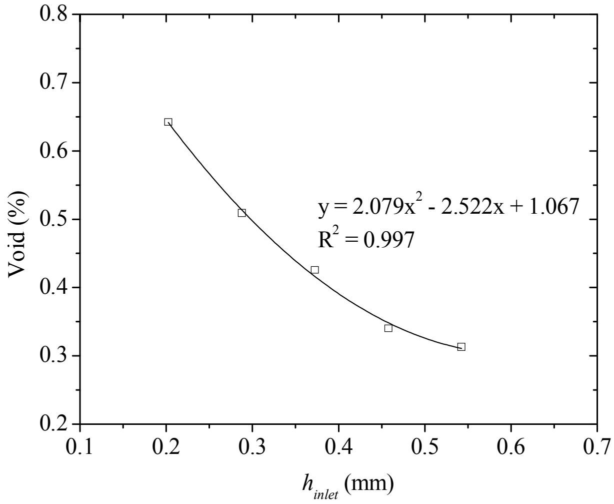


Figure 15. Percentage of void formation in the packages

The positions of the void formation in the packages are presented in Figure 16. The voids are observed at the bottom gap (h_2) for all cases because of the smaller gap between the mould and chip surface. Figure 16 shows the shape and position of the voids, which are concentrated at the edge of the chips nearer to the outlet. Almost identical locations were identified in all cases with only small difference in void shape. The void formation phenomenon was investigated and the flow of the EMC is illustrated in Figure 17. The faster flow in the free regions and gap h_1 completely covers the stacking chips while the air is trapped in the space between the chips and the bottom mould wall, making the EMC unable to flow through the unfilled region. Thus, a void is formed in the package.

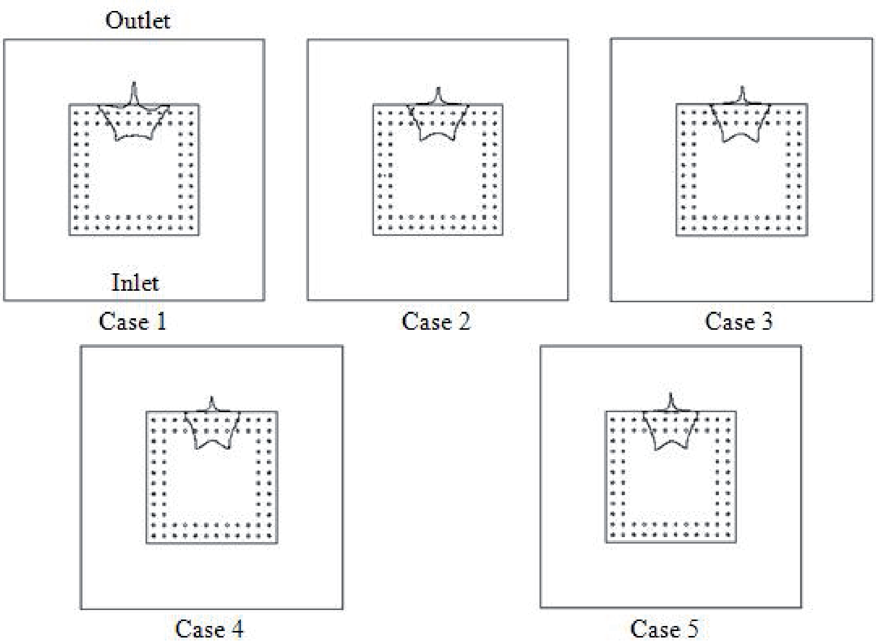


Figure 16. Location of voids from the bottom view

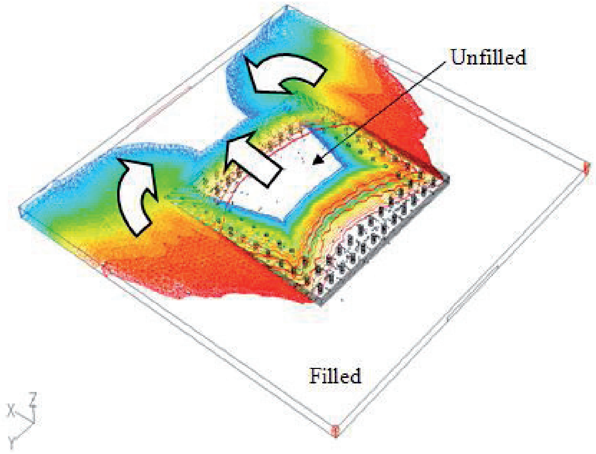


Figure 17. Phenomenon of void formation

Drag Force on Stacking Chips and TSVs

Viscous fluid flow induces drag force to the silicon chips and TSVs during the encapsulation process. The increase in h_{inlet} increases the drag force of the EMC, as shown in Figure 18. The drag force is directly proportional to h_{inlet}. The force acting on the stacking chips and TSVs can cause undesirable displacement and the tendency of the TSVs and chips to fracture in subsequent processes.

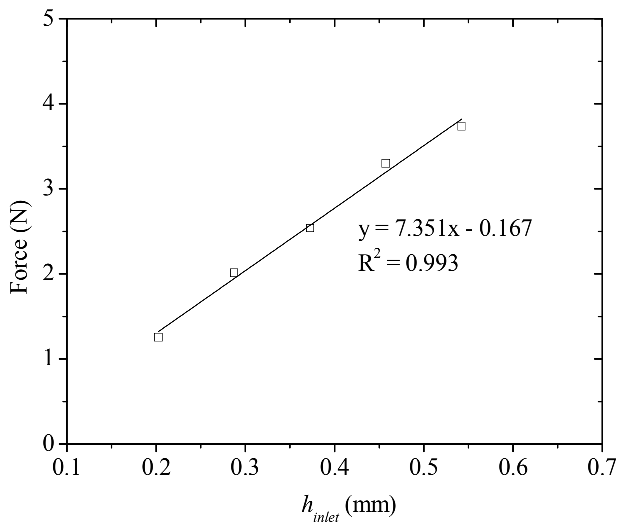


Figure 18. Drag force acting on the stacking chips and TSVs at time 2.6 sec.

Structural Analysis

The drag force induced by the viscous fluid flow was considered in the structural analysis. The TSV structure was modelled and analysed using the finite-element-based software ABAQUS 6.9. Figure 19 shows a schematic diagram of the fluid flow through the silicon chip and TSVs. In the present study, the simplification of the copper TSV integration is considered in the analysis and the force acting on the structure is assumed as uniform in the *x*-direction corresponding to the direction of the EMC flow. From the analysis, the displacement of the TSVs in the *x*-direction as a function of h_{inlet} is plotted in Figure 20; an increase in the stacking chips causes a higher displacement and consequent reduction in their rigidity. Figure 21 shows the displacement profile of the TSV structure with different numbers of stacking chips. The displacement occurs mainly at the top of the structure. This situation may be attributed to the fixed boundary condition at the bottom of the TSVs and the fluid-structure interaction that subjects the TSVs to the drag force of the fluid flow. The drag force induced by the EMC fluid deforms the TSV structure and the von Mises stress [21-23] is used to evaluate the stress on the structure. The von Mises stress of the TSV structure for each case is presented in Figure 22, which illustrates the von Mises stress distribution on the TSVs. The stress distribution shows a quadratic relationship with the increasing number of stacking chips. The stress is concentrated at the first layer of the TSV, which is near the fixed boundary as illustrated in Figure 23. A proper package design and control of the encapsulation process is important in minimising the structure deformation and stress of TSVs to maintain package reliability.

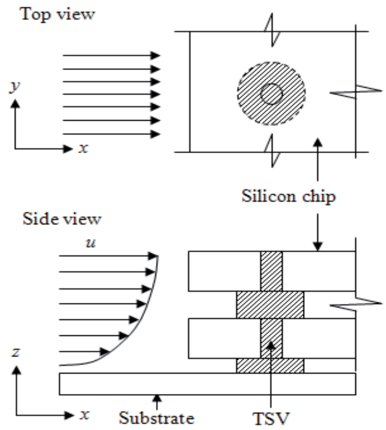


Figure 19. Fluid flow through silicon chips and TSVs

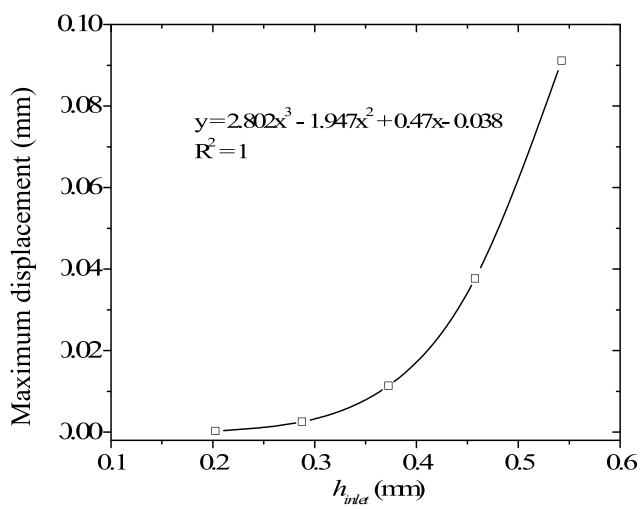


Figure 20. Maximum displacement of TSVs in x-direction

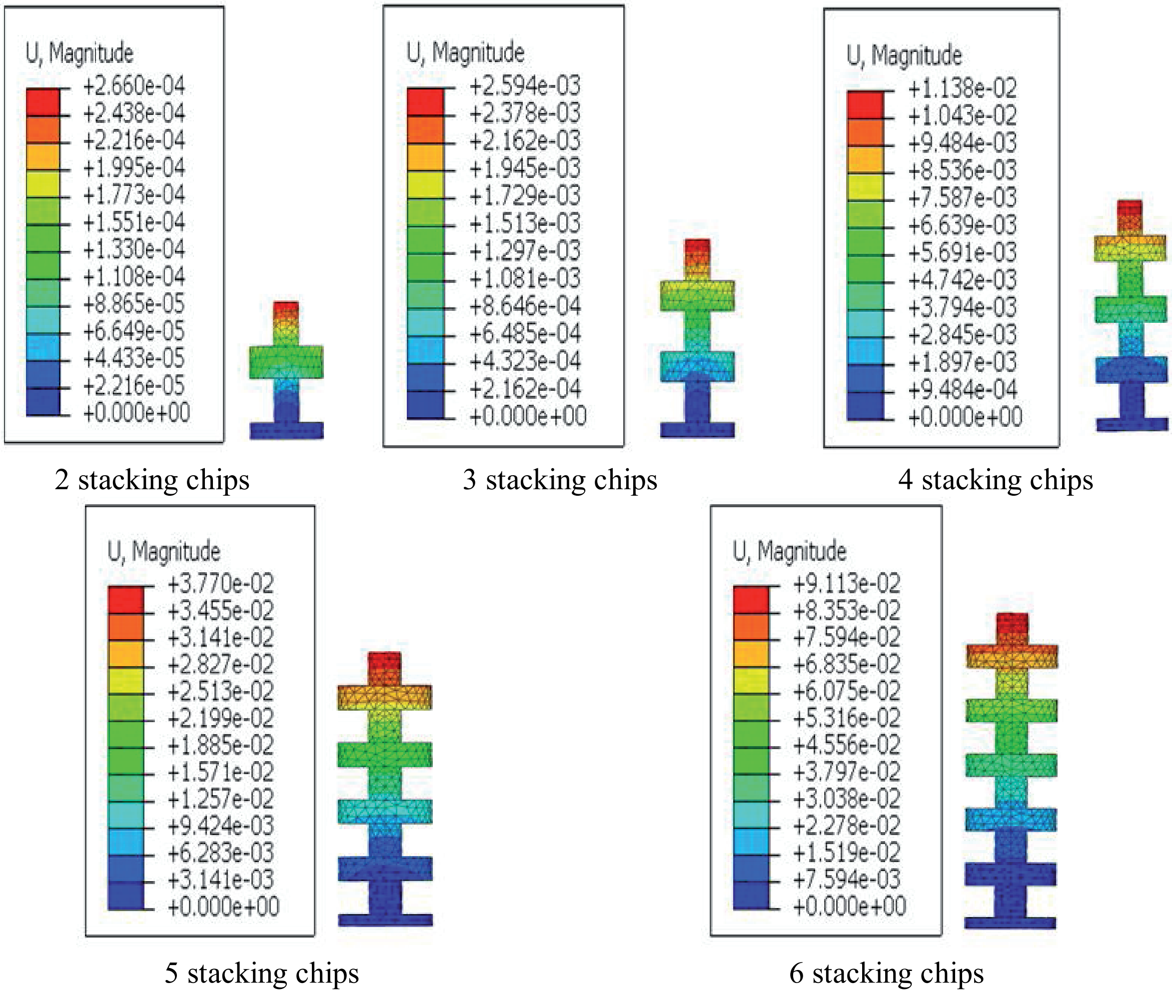


Figure 21. Magnitude of displacement (U, mm) profile of TSVs with different numbers of stacking chips

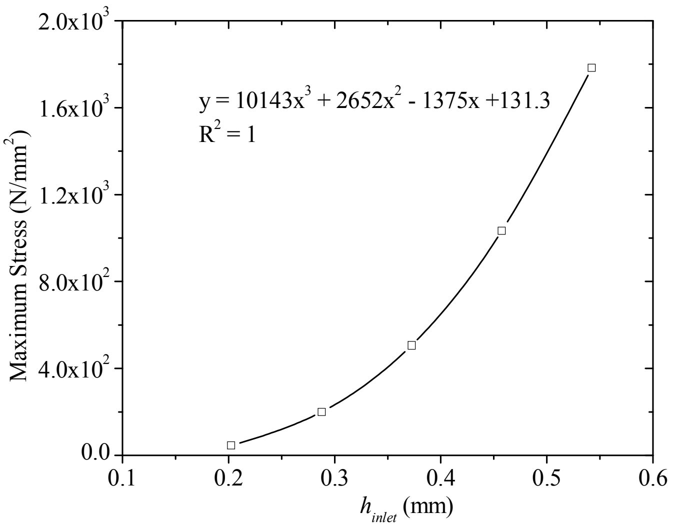


Figure 22. Maximum von Mises stress on TSV

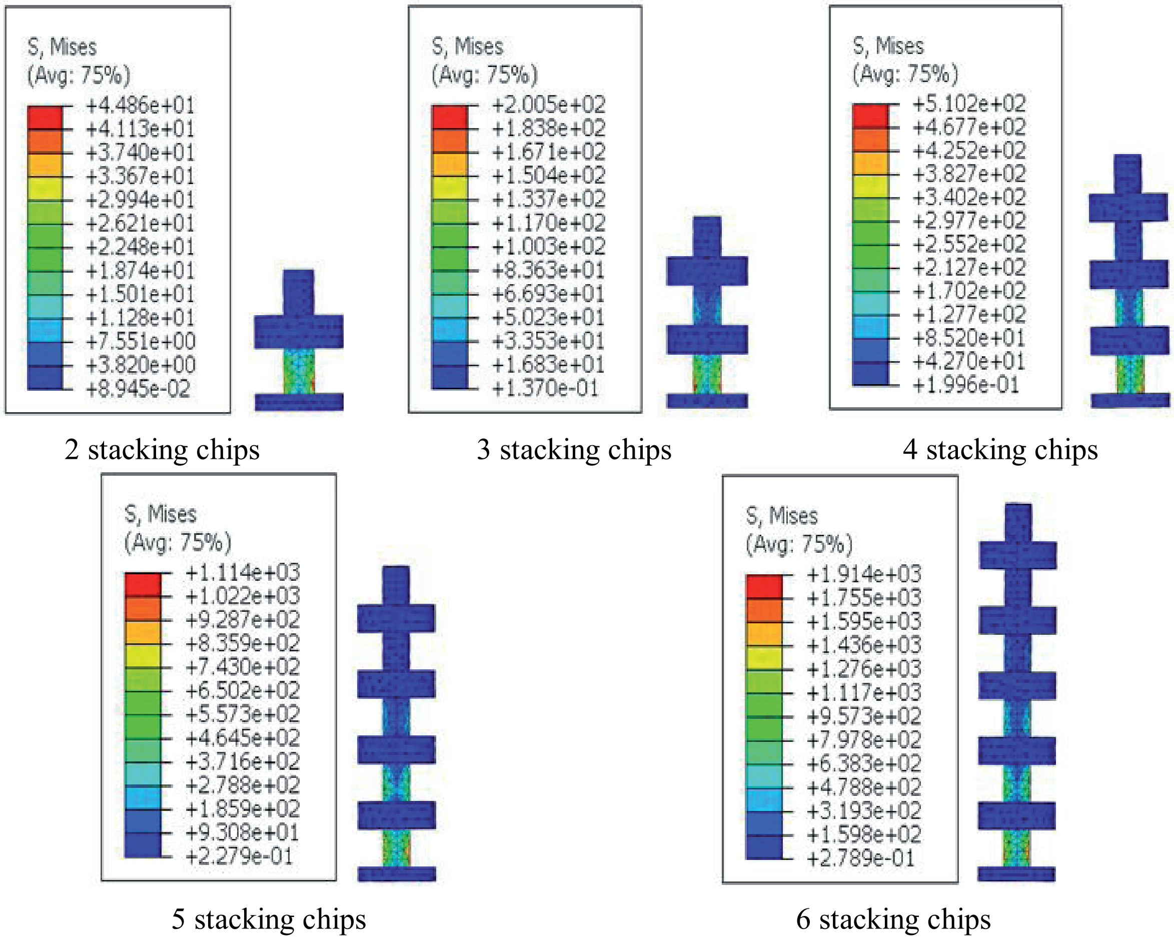


Figure 23. Von Mises stress (N/mm^2) profile of TSVs with different numbers of stacking chips

CONCLUSIONS

The application of the designed inlet-outlet heights in the encapsulation of 3D SCP has been investigated using FLUENT and ABAQUS softwares. The computational time indicated a direct correlation to the number of meshing elements in the finite volume method modelling. Increasing the inlet gate height in the IC packaging maintained the encapsulation time and reduced the percentage of void formation by slightly influencing the void shape in the package as the number of stacking chips increased. The void location was identified to be nearly identical in all cases at gap h_2 close to the outlet gate. It was found that the pressure and drag force induced by the EMC, which caused a larger displacement at the top part of the TSVs was directly proportional to the inlet gate height. However, the stress was concentrated at the lower part of the TSVs.

ACKNOWLEDGEMENTS

The authors would like to thank the Ministry of Higher Education, Malaysia and Universiti Sains Malaysia for a Ph.D. scholarship and financial support for this research work.

REFERENCES

1. J. H. Lau, "TSV manufacturing yield and hidden costs for 3D IC integration", Proceedings of Electronic Components and Technology Conference, 1-4 June **2010**, Las Vegas (NV), USA, pp.1031-1042.
2. K. N. Tu, "Reliability challenges in 3D IC packaging technology", *Microelectron. Reliabil.*, **2011**, *51*, 517-523.
3. R. Han, L. Shi and M. Gupta, "Three-dimensional simulation of microchip encapsulation process", *Polym. Eng. Sc.*, **2000**, *40*, 776-785.
4. S. Moon, C. Yang, Z. Li and A. Fischer, "Three dimensional flow analysis during injection moulding process for stacked-die packages", Proceedings of IEEE/CPMT International Electronic Manufacturing Technology Symposium, 3-5 Oct. **2007**, San Jose (CA), USA, pp.20-23.
5. L. Nguyen, C. Quentin, W. Lee, S. Bayyuk, S. A. Bidstrup-Allen and S.-T. Wang, "Computational modelling and validation of the encapsulation of plastic packages by transfer moulding", *J. Electron. Packag.*, **2000**, *122*, 138-146.
6. C. Y. Khor, M. Z. Abdullah, M. K. Abdullah, M. A. Mujeebu, D. Ramdan, M. F. M. A. Majid, Z. M. Ariff and M. R. A. Rahman, "Numerical analysis on the effects of different inlet gates and gap heights in TQFP encapsulation process", *Int. J. Heat Mass Transf.*, **2011**, *54*, 1861-1870.
7. R. Y. Chang, W. H. Yang, E. Chen, C. Lin and C. H. Hsu, "On the dynamics of air-trap in the encapsulation process of microelectronic package," Proceedings of Annual Technical Conference, 26-30 April **1998**, Atlanta (GA), USA, pp.1178-1180.
8. R. Y. Chang, W. H. Yang, S. J. Hwang and F. Su, "Three-dimensional modeling of mold filling in microelectronics encapsulation process", *IEEE Trans. Compon. Packag. Technol.*, **2004**, *27*, 200-209.
9. M. K. Abdullah, M. Z. Abdullah, S. Kamaruddin and Z. M. Ariff, "Study of flow visualization in stacked-chip scale package (S-CSP)", *Int. Comm. Heat Mass Transfer*, **2007**, *34*, 820-828.
10. T. Schreier-Alt, F. Rehme, F. Ansorge and H. Reichi, "Simulation and experimental analysis of large area substrate overmolding with epoxy molding compounds", *Microelectron. Reliabil.*, **2011**, *51*, 668-675.
11. M. W. Lee, W. K. Jung, E. S. Sohn, J. Y. Lee, C. H. Hwang and C. H. Lee, "A study on the rheological characterization and flow modeling of molded underfill (MUF) for optimized void elimination design", Proceedings of Electronic Components and Technology Conference, 27-30 May **2008**, Lake Buena Vista (FL), pp.382-388.
12. C-T. Ko and K-N. Chen, "Wafer-level bonding/stacked technology for 3D integration", *Microelectron. Reliabil.*, **2010**, *50*, 481-488.
13. T. Falat, K. Friedel, N. Marenco and S. Warnat, "TSV constraints related to temperature excursion, pressure during molding, materials used and handling loads", *Microsyst. Technol.*, **2009**, *15*, 181-190.

14. F. Wang, X. Li, R. Cheng, K. Jiang and S. Feng, "Silicon cantilever arrays with by-pass metal through-silicon-via (TSV) tips for micromachined IC testing probe cards", *Microelectron. Eng.*, **2009**, *86*, 2211-2216.
15. X. Gagnard and T. Mourier, "Through silicon via: From the CMOS imager sensor wafer level package to the 3D integration", *Microelectron. Eng.*, **2010**, *87*, 470-476.
16. M. Abouelatta-Ebrahim, R. Dahmani, O. Valorge, F. Calmon and C. Gontrand, "Modelling of through silicon via and devices electromagnetic coupling", *Microelectron. J.*, **2011**, *42*, 316-324.
17. T. Ohba, N. Meada, H. Kitada, K. Fujimoto, K. Suzuki, T. Nakamura, A. Kawai and K. Arai, "Thinned wafer multi-stack 3DI technology", *Microelectron. Eng.*, **2010**, *87*, 485-490.
18. J. N. Burghartz, W. Appel, C. Harendt, H. Rempp, H. Richter and M. Zimmermann, "Ultra-thin chip technology and applications, a new paradigm in silicon technology", *Solid-State Electron.*, **2010**, *54*, 818-829.
19. C. J. Wu, M. C. Hsieh, C. C. Chiu, M. C. Yew and K. N. Chiang, "Interfacial delamination investigation between copper bumps in 3D chip stacking package by using modified virtual crack closure technique", *Microelectron. Eng.*, **2011**, *88*, 739-744.
20. L. J. Ladani, "Numerical analysis of thermo-mechanical reliability of through silicon vias (TSVs) and solder interconnects in 3-dimensional integrated circuits", *Microelectron. Eng.*, **2010**, *87*, 208-215.
21. L. J. Ladani, "Stress analysis of 3-dimensional IC package as function of structural design parameters", *Microelectron. Eng.*, **2010**, *87*, 1852-1860.
22. A. Kamto, Y. liu, L. Schaper and S. L. Burkett, "Reliability study of through-silicon via (TSV) copper filled interconnects", *Thin Solid Films*, **2009**, *518*, 1614-1619.
23. L-C. Shen, C-W. Chien, H-C. Cheng and C-T. Lin, "Development of three-dimensional chip stacked technology using a clamped through-silicon via interconnection", *Microelectron. Reliabil.*, **2010**, *50*, 489-497.
24. Y. C. Tan, C. M. Tan, X. W. Zhang, T. C. Chai and D. Q. Yu, "Electromigration performance of through silicon via (TSV)- A modeling approach", *Microelectron. Reliabil.*, **2010**, *50*, 1336-1340.
25. C. J. Wu, M. C. Hsieh and K. N. Chiang, "Strength evaluation of silicon die for 3D chip stacked packages using ABF as dielectric and barrier layer in through-silicon via", *Microelectron. Eng.*, **2010**, *87*, 505-509.
26. M. Y. Tsai and C. H. Chen, "Evaluation of test methods for silicon die strength", *Microelectron. Reliabil.*, **2008**, *48*, 933-941.
27. C. P. Wong and Z. Zhang, "Recent advances in flip-chip underfill: Materials, process, and reliability", *IEEE Trans. Adv. Packag.*, **2004**, *27*, 515-524.
28. P. Alpern, R. Dudek, R. Schmidt, V. Wicher and R. Tilgner, "On the mode II popcorn effect in thin packages", *IEEE Trans. Compon. Packag. Technol.*, **2002**, *25*, 56- 65.
29. S.-Y. Teng and S.-J. Hwang, "Simulations and experiments of three-dimensional paddle shift for IC packaging", *Microelectron. Eng.*, **2008**, *85*, 115-125.

30. C. C. Pei and S. J. Hwang, “Three-dimensional paddle shift modeling for IC packaging”, *J. Electron. Packag.*, **2005**, *127*, 324-334.
31. C. Y. Khor, M. Z. Abdullah and F. Che Ani, “Study on the fluid/structure interaction at different inlet pressures in moulded packaging”, *Microelectron. Eng.*, **2011**, *88*, 3182-3194.
32. R. Lannuzzelli, “Predicting plated-through-hole reliability in high temperature manufacturing processes”, Proceedings of 40th Electronic Components and Technology Conference, 11-16 May **1991**, Atlanta (GA), USA, pp.410-421.
33. C. Y. Khor, M. K. Abdullah, M. Z. Abdullah, M. Abdul Mujeebu, D. Ramdan, M. F. M. A. Majid and Z. M. Ariff, “Effect of vertical stacking dies on flow behaviour of epoxy moulding compound during encapsulation of stacked-chip scale packages”, *Heat Mass Transf.*, **2010**, *46*, 1315-1325.
34. C. Y. Khor, Z. M. Ariff, F. Che Ani, M. Abdul Mujeebu, M. K. Abdullah, M. Z. Abdullah and M. A. Joseph, “Three-dimensional numerical and experimental investigations on polymer rheology in meso-scale injection moulding”, *Int. Comm. Heat Mass Transf.*, **2010**, *37*, 131-139.
35. C. Nitatwichit, Y. Khunatorn and N. Tippayawong, “Computational analysis and visualisation of wind-driven naturally ventilated flows around a school building”, *Maejo Int. J. Sci. Technol.*, **2008,** *2*, 240-254.

Maejo Int. J. Sci. Technol. **2012**, *6*(02), 186-195

Maejo International
Journal of Science and Technology

ISSN 1905-7873
Available online at www.mijst.mju.ac.th

Full Paper

Biochar production from freshwater algae by slow pyrolysis

Kanyaporn Chaiwong [1], Tanongkiat Kiatsiriroat [1, *], Nat Vorayos [1] and Churat Thararax [2]

[1] Department of Mechanical Engineering, ChiangMai University, Chiang Mai50200, Thailand
[2] Department of Mechanical Engineering, Rajamangala University of Technology Lanna, Chiang Mai 50300, Thailand

* Corresponding author, e-mail: kiatsiriroat_t@yahoo.co.th

Received: 8 August 2011 / Accepted: 8 May 2012 / Published: 11 May 2012

Abstract: A study on the feasibility of biochar production from 3 kinds of freshwater algae, viz. Spirulina, Spirogyra and Cladophora, was undertaken. Using a slow pyrolysis process in a specially designed reactor, biochar could be generated at 550^{o}C under nitrogen atmosphere. The yields of biochar were between 28-31% of the dry algae.

Keywords: biochar, slow pyrolysis, carbon-negative technology, algae, Spirulina, Spirogyra, Cladophora

---

INTRODUCTION

Biochar technology is promoted as a carbon-negative technology [1]. Showing high potential as a solid biofuel, biochar can also be used for carbon sequestration and for improving soil conditions [2]. Biochar is a stable form of carbon [1-7], being more stable than the organic form and capable of remaining in the soil for hundreds or thousands of years [1-2]. This means that biochar can perform an important role in helping to sequester carbon from the atmosphere [6-8]. A considerable reduction in nitrous oxide (N_2O) emissions from biochar-amended soils have also shown [1-2]. N_2O is a greenhouse gas 300 times more potent than CO_2 and it is mainly associated with the use of nitrogen fertilisers [9]. Being highly porous, the biochar structure also helps improve soil conditions and water retention capability of the soil as well as increase its nutrients and surface area [10].

Biochar is produced from a thermochemical process in which biological materials (wood and wood waste, energy crops, aquatic plants and their waste by-products) are heated under a condition of limited or no oxygen [8]. The thermochemical conversion can be categorised into gasification, pyrolysis and liquefaction [11]. The pyrolysis process is usually used for biochar production because the technique is relatively simple and inexpensive and allows considerable flexibility in both the type and quality of the biomass feedstock [12]. Pyrolysis can produce the biofuel as gas, bio-oil and biochar. The proportion of the products is dependent on the process, which can be divided into 3 subclasses: slow pyrolysis, fast pyrolysis and intermediate pyrolysis [13], depending on the operating conditions. The proportions of the produced forms of fuel from wood are shown in Table 1 [14]. The slow pyrolysis, which gives a maximum yield of biochar, has been widely used for decades in charcoal kilns where the combustion occurs in the absence of oxygen. A slow heating rate is maintained up to 400-700^{o}C [15]. At this low temperature range, a high carbon recovery from the organic biomass is obtained in the pyrolised biochar. For a higher temperature range of 700-800^{o}C, the cation exchange capacity of biochar itself can also be improved, but a lower carbon yield (about 5% loss) is obtained. The optimum biochar production temperature in terms of carbon is recommended at 500^{o}C [16].

Table 1. Mean composition of post-pyrolysis feedstock (wood) residues from different pyrolytic conditions [14]

Process	Liquid (bio-oil)	Solid (biochar)	Gas (syngas)
Fast Pyrolysis Moderate temperature (~500 °C), short hot-vapour residence time (<2s)	75% (25% water)	12%	13%
Intermediate Pyrolysis Low-moderate temperature, moderate hot-vapour residence time	50% (50% water)	25%	25%
Slow Pyrolysis Low-moderate temperature, long residence time	30% (70% water)	35%	35%

Algae have been promoted as a source of energy due to their advantages such as their fast growth rate, which is 10-340 times that of oil crops, and their less requirement of cultivation area. Besides, algae cultivation can be carried out not only in natural waters, i.e. sea water and freshwater, but also under such poor condition as in wastewater. This means that the algae can produce bio-energy without competition with the food production sector for cultivation space [17]. The energy conversion from algae can be carried out by 2 processes [18]. One is the biochemical conversion by which biofuel in the forms of biodiesel or ethanol such as that from marine green alga *Chlorococcum littorale* by dark fermentation [19] can be produced. For biodiesel, a specific type of algae, such as *Botryococcus braunii* or *Chlorella* spp., which give a high oil content of 30-75% wt of dry algae is normally required [20]. The second process is the thermochemical conversion that changes all types of algae into energy products via heat reaction; biochar is a by-product from this process. Algal biochar has a lower carbon content, surface area and cation exchange capacity compared with the lignocellulose biochar but has a higher pH and gives a higher content of nitrogen, ash and inorganic elements (P, K, Ca and Mg). The algal biochar could therefore be utilised to reduce the acidity of the soil and increase its inorganic nutrients [21]. However, only a few studies have been reported on biochar production from algae.

In this paper, a study on algal biochar production from three types of freshwater algae, viz. Spirulina (*Spirulina* spp.), Spirogyra (*Spirogyra* spp.) and Cladophora (*Cladophora* spp.), was carried out by slow pyrolysis technique. These algae, compared to other algae types used for biodiesel production [18], have a low oil content.

MATERIALS AND METHODS

Algae and Preliminary Analysis

Three studied samples of dry freshwater algae, viz. Spirulina (*Spirulina* spp.), Spirogyra (*Spirogyra* spp.) and Cladophora (*Cladophora* spp.), were of commercial grade (Figure 1). The first one is the micro-type and the other two are the macro-type. Spirulina is a genus of cyanobacteria and the dried samples are available in the local market. Spirogyra (locally known as Tao) and Cladophora (locally known as Kai) are common filamentous green algae [22] which can be found and collected from Nan River in the northern part of Thailand.

Before pyrolysis, each algal sample was characterised using ultimate analysis and proximate analysis. The former was performed as described by Channiwala and Parikh [23] while the latter was performed as described by Parikha et al [24].

Figure 1. Samples of dry algae: (a) Spirulina; (*b*) Cladophora; (c) Spirogyra

Thermo-Decomposition of Algae

Thermo-decomposition of organic matter under pyrolysis can be characterised by thermogravimetric analysis (TGA), in which the amount and rate of change in the weight of material due to decomposition, oxidation or dehydration is measured as a function of temperature in the form of a TG curve. Its derivative with respect to temperature is a DTG curve. The temperature at the peak (minimum or maximum) of the DTG curve shows the activation of the thermochemical reaction. If the peak of the DTG curve occurs at a low temperature, the reaction can be easily performed while the height of the curve corresponds to the capability to release volatile matter during the volatilisation reaction.

In this study, the tested algae were fed into a thermogravimetric analyser (PerkinElmer TGA7) and the tested conditions were controlled under nitrogen atmosphere. The heating program was: 50-135^{o}C (10^{o}/min.); constant at a 135^{o}C (5 min.); 135-900^{o}C (100^{o}/min). After that, the temperature was kept constant under oxygen atmosphere for 15 min. The TG and DTG plots could then be performed from the experimental data.

Fixed-Bed Pyrolysis

A 125-g sample of each dried alga was fed into a stainless steel fixed-bed reactor (21 cm high and 6 cm in diameter). The experimental set-up is shown in Figure 2. Nitrogen gas was fed at a flow rate of 30 ml/min. for 30 min. to remove air in the reactor before heating. The reactor was heated up at a rate of 8^{o}C/min. until the temperature reached a set temperature of 550^{o}C, at which it was then kept constant for 60 min. The gas leaving the reactor was condensed in two water-cooled condensers and the liquid (bio-oil) was stored in two collection flasks while the solid residue (biochar) remained in the reactor. Each experiment was performed in triplicate and the results averaged. After pyrolysis, the biochar products were subjected to ultimate and proximate analyses in the same manner as the dry algae before pyrolysis.

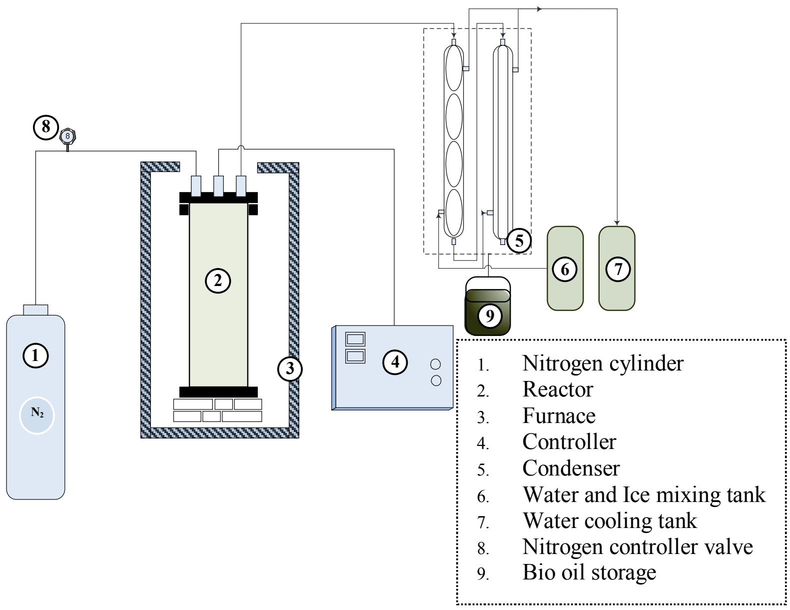


Figure 2. A schematic sketch of the pyrolysis experimental set-up

RESULTS AND DISCUSSION

The results of ultimate and proximate analyses of the dried algae are shown in Table 2. While Spirogyra and Cladophora are macrophytic green algae and Spirulina is a microphytic blue-green algae, the composition characteristics of Spirulina are quite similar to those of Spirogyra. For instance, both have a high carbon content and HHV, which should make them particularly suitable as raw material for generating biochar.

The pyrolytic characteristics of the three algae at temperature between 50-900^{o}C as determined by TGA are shown in Figure 3. The percentage of the existing weight as function of temperature (TG curve) showed that there were three stages occurring during the pyrolysis process. The first stage (up to about 200^{o}C) was dehydration; the weight loss was due to moisture removal from the sample. The next stage (200-600°C) was devolatilisation; the weight loss in this stage was about 60% due to the loss of volatile components. The last stage (600-900°C) was solid decomposition; at this stage the weight loss was slower. The DTG plots of the tested algae are also shown in Figure 3. In the devolatilisaton stage between 200-600°C, the peak in the DTG plot for Spirulina was highest compared with Spyrogra and Cladophora, which means that highest volatile matter was released resulting in highest yield of bio-oil.

Table 2. Results of ultimate and proximate analyses of dried algae

Analysis	Algae		
	Spirogyra	Cladophora	Spirulina
Ultimate Analysis (%)			
Sulfur	0.57	5.29	0.49
Carbon	39.26	28.78	42.83
Hydrogen	6.11	4.02	6.02
Nitrogen	6.65	3.06	4.09
Oxygen*	47.41	58.85	46.57
Proximate Analysis (%)			
Moisture	8.45	5.00	8.04
Ash	13.99	33.42	6.98
Volatiles	65.48	60.61	68.15
Fixed Carbon	12.08	0.98	16.83
HHV (MJ/kg) **	22.34	16.22	23.42

* By difference; ** Higher heating value, estimated by correlation of Channiwala and Parikh [23]

Normally, organic material is mainly composed of cellulose, hemicelluloses, lignin and extractive matter. Due to a lower cellulose and hemicellulose content and higher extractive matter content in algae compared with other kinds of biomass, a lower pyrolysis reaction temperature of the former can be expected. Table 3 compares the decomposition of algae and some other kinds of biomass. The temperatures at which the highest amount of volatiles in the biomass could be expelled out (temperatures at maximum decomposition rate) are given. It can be seen that the temperatures for the algae were around 300-330°C, which were lower than those for other forms of biomass. Thus, the pyrolysis reaction of the algae could occur more easily and with lower energy input. The maximum decomposition rate was also lower, which indicates that the amounts of volatile matter from the algae or the yields of bio-oil were lower. On the other hand, the amounts of solid residue from algae (biochar and ashes) at the end of the pyrolysis were observed to be higher than those from other forms of biomass.

The products and their distribution resulting from the slow pyrolysis of the three algae in the specially designed reactor are given in Figure 4 and Table 4 respectively, from which it can be observed that the bio-oil yield from Spirulina was highest and that from Cladophora was lowest. The biochar yields from all algae samples were somewhat similar, viz. 28-31% by weight of the dry algae.

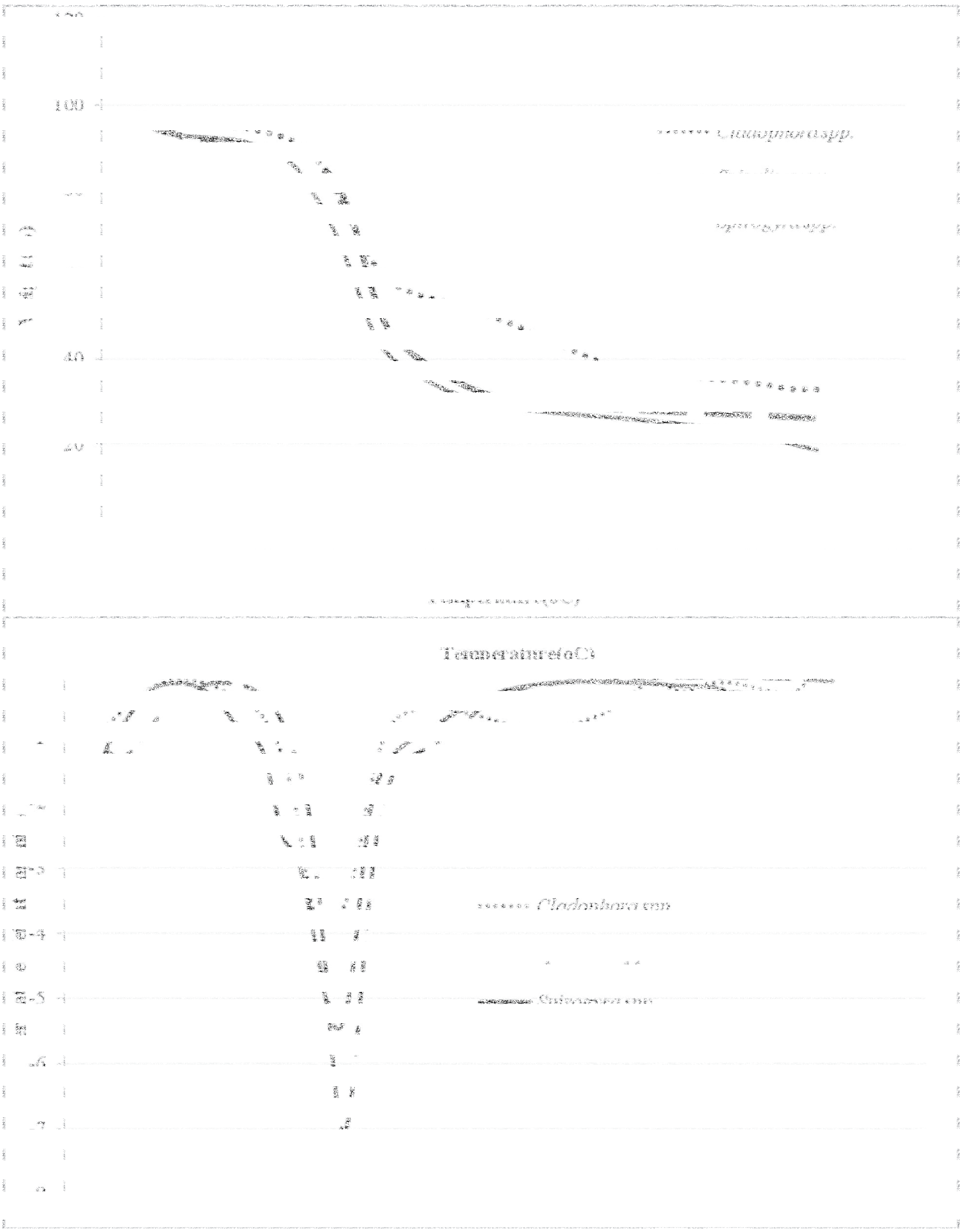


Figure 3. TG (top) and DTG (bottom) plots for tested algae

Table 3. Comparison of decomposition of algae and other forms of biomass [25]

Biomass	Temperature at maximum decomposition rate (°C)	Maximum decomposition rate (% min^{-1})**	% Solid residue at 900°C
Corncob	346	27.9	10.00
Rice husk	367	31.0	12.60
Eucalyptus	361	31.0	13.30
Sawdust	367	18.6	13.60
Palm shell	377	17.9	20.80
Coconut shell	355	18.2	22.07
Spirulina*	324	7.04	25.96
Spirogyra*	318	5.49	18.61
Cladophora*	300	3.92	32.21

* Results from present work
** Value of maximum decomposition rate is the percentage of maximum weight loss rate from DTG compared with initial weight.

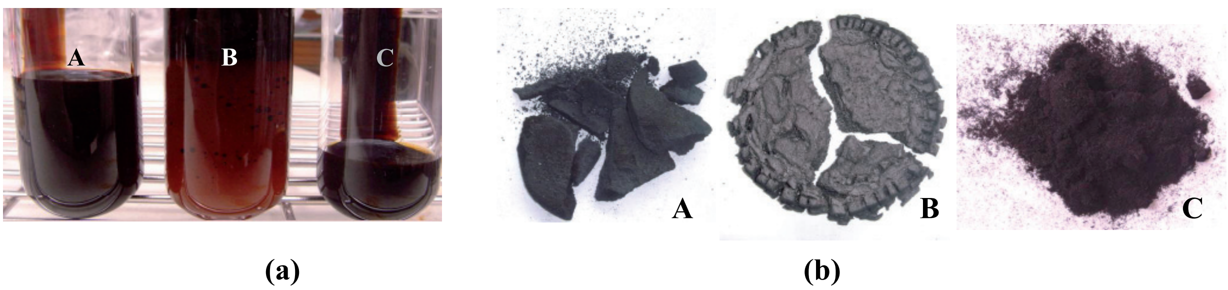


Figure 4. Bio-oil and biochar from slow pyrolysis of algae (A: Spirogyra, B: Spirulina, C: Cladophora)

Table 4. Product distribution from pyrolysis of algae at 550°C

Product	Product distribution (%wt)		
	Spirulina	Spirogyra	Cladophora
Gas*	23	29	30
Liquid (bio-oil)	46	43	39
Solid residue (biochar)	31	28	31

* By difference

The results of proximate and ultimate analyses of algal biochar products and other solid fuels are shown in Table 5. A high-quality solid fuel should have a high amount of fixed carbon with low volatile components and ash content or its chemical components should consist of a high carbon and hydrogen content in order to give a high heating value [26]. As shown in Table 5, peat (S-H3), German Braunkohle lignite and charcoal represent solid fossil fuels from low to high quality. Biochar from Spirulina and Spirogyra have the fixed carbon content similar to that of German Braunkohle lignite while their volatile components are lower. Spirogyra biochar also has the carbon content close to that of German Braunkohle lignite, although its hydrogen content is lower and nitrogen content higher. Thus, biochar from Spirulina and Spirogyra has high potential, after ash removal, to be used as fuel.

Table 5. Results of proximate and ultimate analyses of algal biochar compared with other solid fuels

Name	Fixed carbon %	Volatiles %	Ash %	HHV (MJ/kg)	C %	H %	O %	N %	S %	Ref.
Peat (S-H3)	26.87	70.13	3.00	22.00	54.81	5.38	35.81	0.89	0.11	[26]
German Braunkohle lignite	46.03	49.47	4.50	25.10	63.89	4.97	24.54	0.57	0.48	[24]
Charcoal	89.10	9.88	1.02	34.39	92.04	2.45	2.96	0.53	1.00	[24]
Oak char	59.30	25.80	14.90	24.80	67.70	2.40	14.40	0.40	0.20	[24]
Cladophora coelothrix char	-	-	32.10	-	34.60	1.50	-	3.30	-	[21]
Cladophora patentiramea char	-	-	47.00	-	20.30	1.20	-	1.70	-	[21]
Spirulina char	44.55	7.63	47.82	15.78	45.26	1.24	0.28	2.57	0.07	*
Cladophora char	26.68	35.50	37.81	16.68	51.14	0.56	0.69	1.98	1.86	*
Spirogyra char	59.66	16.81	23.53	22.96	62.37	0.37	4.07	2.11	0.48	*

* Results from present work

Lehmann et al. [8] suggested that the efficiency of carbon sequestration by algal bio-char is obtained when the carbon conversion into biochar leads to a sequestration of about 50% of the initial carbon. In this study, as seen from Table 5, the biochar from Spirogyra seems to be the most suitable candidate for carbon sequestration. The carbon content of Spirogyra char is also close to that of oak char and higher than that of the recently studied green tide algae (*Cladophora coelothrix* and *C. patentiramea* [21]). The high ash and nitrogen content in the algal biochar can also be beneficial to agricultural soil, especially acidic or nutrient-defficient soil [10].

CONCLUSIONS

This study has demonstrated an algal biochar production process by slow pyrolysis for three types of freshwater algae, namely Spirulina, Spirogyra and Cladophora. The quality of the biochar obtained was comparable to that of certain established solid fuels as well as biochar obtained from other forms of biomass. However, the temperature used for production appeared to be lower.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Joint Ph.D. Program under the Strategic Scholarships for Frontier Research Network by the Office of the Higher Education Commission (Thailand) and the financial support from the National Research University Project. The Graduate School and Department of Mechanical Engineering, Chiang Mai University are thanked for technical support. The authors also thank Professor Dr. Joachim Müller from Hohenheim University, Germany for the valuable guidance on this study.

REFERENCES

1. J. A. Mathews, "Carbon-negative biofuels", *Energy Policy*, **2008**, *36*, 940-945.
2. J. Lehmann and S. Joseph (Ed.), "Biochar for environmental management: Science and Technology", Earthscan, London, **2009**, pp.1-9.
3. L. C. R. Pessenda, S. E. M. Gouveia and R. Aravena, "Radiocarbon dating of total soil organic matter and humin fraction and its comparison with ^{14}C ages of fossil charcoal", *Radiocarbon,* **2001**, *43,* 595-601.
4. M. W. I. Schmidt, J. O. Skjemstad and C. Jager, "Carbon isotope geochemistry and nanomorphology of soil black carbon: Black chernozemic soils in central Europe originate from ancient biomass burning", *Glob. Biogeochem. Cycles*, **2002**, *16*, 1123-1130.
5. E. S. Krull, C. W. Swanston, J. O. Skjemstad and J. A. McGowan, "Importance of charcoal in determining the age and chemistry of organic carbon in surface soils", *J. Geophys. Res.*, **2006,** *111,* 4001-4009.
6. M. J. Antal, E. Croiset, X. Dai, C. DeAlmeida, W. S. L. Mok, N. Norberg, J.-R. Richard and M. Majthoub, "High-yield biomass charcoal", *Ener. Fuels*, **1996,** *10,* 652-658.
7. B. D. Okello, T. G. O'Connor and T. P. Young, "Growth, biomass estimates, and charcoal production of *Acacia drepanolobium* in Laikipia, Kenya", *For. Ecol. Manage.*, **2001**, *142*, 143-153.
8. J. Lehmann, J. Gaunt and M. Rondon, "Bio-char sequestration in terrestrial ecosystems–A review", *Mitig. Adapt. Strat. Glob. Change,* **2006**, *11,* 403-427.
9. Y. Yanai, K. Toyota and M. Okazaki, "Effects of charcoal addition on N_2O emissions from soil resulting from rewetting air-dried soil in short-term laboratory experiments", *Soil Sci. Plant Nutr.*, **2007**, *53*, 181-188.
10. S. Sohi, E. Lopez-Capel, E. Krull and R. Bol, "Biochar, climate change and soil: A review to guide future research", CSIRO, Clayton South (Australia), **2008**.
11. P. McKendry, "Energy production from biomass (Part 2): Conversion technologies", *Bioresour. Technol.*, **2002**, *83,* 47-54.
12. D. A. Laird, R. C. Brown, J. E. Amonette and J. Lehmann, "Review of the pyrolysis platform for coproducing bio-oil and biochar", *Biofuels Bioprod. Biorefin.,* **2009**, *3,* 547-562.

13. P. M. McHenry, "Agricultural bio-char production, renewable energy generation and farm carbon sequestration in Western Australia: Certainty, uncertainty and risk", *Agric. Ecosyst. Environ.*, **2009**, *129,* 1-7.
14. International Energy Agency, "IEA Bioenergy 2006 Annual Report", http://www.ieabioenergy.com/DocSet.aspx?id=5566&ret=lib.
15. H. B. Goyal, D. Seal and R. C. Saxena, "Bio-fuels from thermochemical conversion of renewable resources: A review", *Renew. Sustain. Ener. Rev.*, **2008***, 12*, 504-517.
16. J. Lehmann, "Bio-energy in the black", *Front Ecol. Environ.*, **2007**, *5,* 381-387.
17. Y. Chisti, "Biodiesel from microalgae", *Biotechnol. Adv.*, **2007**, *25,* 294-306.
18. S. Amin, "Review on biofuel oil and gas production processes from microalgae", *Ener. Conver. Manage.*, **2009**, *50*, 1834-1840.
19. Y. Ueno, N. Kurano and S. Miyachi, "Ethanol production by dark fermentation in the marine green alga, *Chlorococcum littorale*", *J. Ferment. Bioeng.*, **1998**, *86*, 38-43.
20. P. Metzger and C. Largeau, "*Botryococcus braunii*: A rich source for hydrocarbons and related ether lipids", *Appl. Microbiol. Biotechnol.*, **2005**, *66,* 486-496.
21. M. Bird, C. M. Wurster, P. H. de P. Silva, A. M. Bass and R. de Nys, "Algal biochar – production and properties"*, Bioresour. Technol.*, **2011**, *102,* 1886-1891.
22. Y. Peerapornpisal, S. Makornkaokeyul, I. Pongsirikul, S. Lertleelakitja, S. Preankarn, S. Theeranupatana and K. Sutjaritwong, "Potential of freshwater macroalgae as food and medicine", Research report to Thailand Research Fund (TRF), Report No. RDG4520031, **2008**, pp.168-170.
23. S. A. Channiwala and P. P. Parikh, "A unified correlation for estimating HHV of solid, liquid and gaseous fuels", *Fuel*, **2002**, *81*, 1051-1063.
24. J. Parikha, S. A. Channiwala and G. K. Ghosal, "A correlation for calculating HHV from proximate analysis of solid fuels", *Fuel*, **2005**, *84*, 487-494.
25. C. Narumon, "Effects of chemical composition of biomass on pyrolysis and combustion", *MS Thesis*, **2007**, Chulalongkorn University*,* Thailand.
26. S. Gaur and T. B. Reed, "Thermal Data for Natural and Synthetic Fuels", Marcel Dekker, New York, **1998**.

Maejo Int. J. Sci. Technol. **2012**, *6*(02), 196-215

Maejo International Journal of Science and Technology

ISSN 1905-7873

Available online at www.mijst.mju.ac.th

Report

Solar technology and building implementation in Malaysia: A national paradigm shift

Syahrul Nizam Kamaruzzaman [1,*], Hamzah Abdul-Rahman [2], Chen Wang [2], Saipol Bari Karim [2] and Tien Yee Lee [1]

[1] Department of Building Surveying, Faculty of Built Environment, University of Malaya, 50603, Kuala Lumpur, Malaysia

[2] Centre for Construction Facility and Project Management, Faculty of Built Environment, University of Malaya, 50603, Kuala Lumpur, Malaysia

* Corresponding author, e-mail: syahrulnizam@um.edu.my

Received: 20 December 2010 / Accepted: 17 May 2012 / Published: 22 May 2012

Abstract: Solar technology is becoming increasingly popular. For example, the production of solar cells quadrupled in the 1999-2004 period, with a capacity of four gigawatts worldwide. Renewable energy including solar power produces few or no harmful emissions and it is becoming increasingly important to exploit it in the future. This paper presents a literature review of the application of numerous types of solar technology in buildings in Malaysia and identifies the challenges faced. Although several newly constructed green buildings use solar technology, Malaysia has yet to accept it wholesale. If solar technology is to be adopted widely, then both public and private sectors must cooperate to provide large-scale financial incentives and produce specialists in solar technology. As the first step, the government has established the Low Energy Office and the Green Energy Office, which use passive solar design and photovoltaic systems in their own buildings. However, the private sector has yet to follow suit. It is anticipated that the application of solar technology in buildings will encourage sustainable development when all non-renewable energy sources decrease significantly. If people do not recognise the potential of such technology in daily life, it will soon be too late.

Keywords: solar technology, photovoltaics, solar energy application, Malaysia, sustainable development.

INTRODUCTION

Solar energy is the most promising source of clean, renewable energy and it has the greatest potential of any power source to solve the world's energy problems [1, 2]. In Malaysia, the climatic conditions are favourable to the development of solar energy, with abundant sunshine throughout the year. The annual average daily solar radiation is 4.21-5.56 kWh/m^2 [3]. The highest solar radiation is estimated at 6.8 kWh/m^2 in August and November, and the lowest is 0.61 kWh/m^2 in December. The northern region and a few places in eastern Malaysia have the highest potential for solar energy application, with particularly high solar radiation throughout the year.

A photovoltaic (PV) system, a source of solar power, consists of several solar cells that convert light energy into electricity [4]. It is an elegant means of producing electricity on site, directly from the sun without concern for fuel supply or environmental impact. Solar power is produced silently with minimum maintenance, no pollution and no resource depletion. The PV system is also exceedingly versatile and can be used to pump water, grind grain and provide communications and village electrification where electricity is unavailable [5]. Currently, most solar energy applications are for domestic hot water systems, water pumping and drying agricultural produce [6]. Most of the solar power used in Malaysia is at the domestic level; it is estimated that over 10,000 units of domestic systems are using the PV system at present while large-scale commercial use is insignificant. Although the PV system has tremendous potential, especially for remote areas in Malaysia, the cost of PV panels and associated technology is extremely high for mass power generation. The current market value of a PV system is about RM 28.00 (US$ 8.40)/Wp [7]. Malaysia has no local solar cell manufacturers and all the PV modules and inverters are imported from countries such as Germany and Japan at a very high cost. As a result, PV systems are not an attractive option for the Malaysian public.

In 1995, a 100-kilowatt peak (kWp) demonstration photovoltaic project was implemented under the initiative of the Ministry of Energy, Water and Communications in Marak Parak, Sabah, heralding a good start for an effective and efficient transfer of PV power generation technology in Malaysia [6]. In order to reduce the cost of the PV systems, in 2005 Malaysia carried out the Malaysia building integrated photovoltaic (MBIPV) project, funded by the government, the global environment facility disbursed through the United Nations development programme (UNDP/GEF), and various private investors [8]. The principal objective of this project was to reduce the long-term cost of building-integrated photovoltaic (BIPV) technology within the Malaysian market, leading to sustainable and widespread BIPV technology applications that avoid greenhouse gas emissions. Over the lifetime of the project, the energy generated will prevent 65,100 tons of CO_2 emissions from the country's power sector [9]. The project is expected to result in a 330% increase in BIPV applications from the 2005 baseline. The full project addresses in an integrated manner the long-term cost reduction of BIPV technology and the adoption of supportive regulatory frameworks to establish the desired environment for a sustainable BIPV market [2]. After the BIPV programme was introduced in 2005, the cost of BIPV dropped to US$ 6500/kWp–a reduction of 40% by 2010 compared to the 2005 cost, as shown in Figure 1.

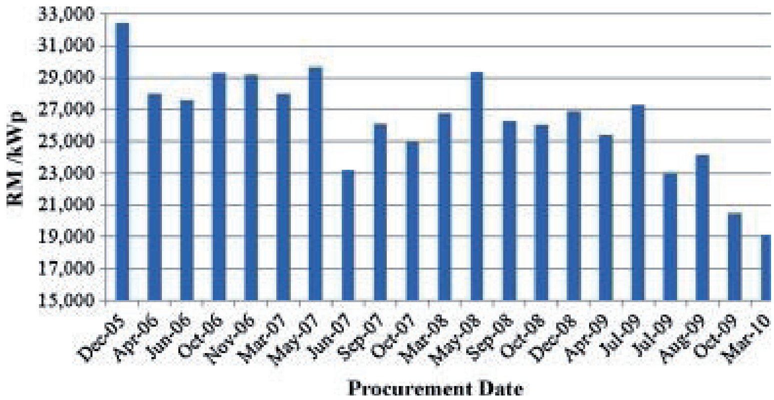


Figure 1. Average BIPV price/kWp from 2005 to 2010 in Malaysia [7]

Another national MBIPV programme, SURIA 1000 [9], which is targeting the residential and commercial sectors, will consolidate the new BIPV market. SURIA 1000 will enable the public and industry to be directly involved in renewable energy and environmental protection initiative. Since 2007, a limited number of grid-connected solar PV systems have been offered to the public through a bidding (auction) system, with a minimum BIPV capacity of 3 kWp per application. Successful bidders install the building-integrated PV system at their premises; the cost of each PV system is borne by the successful bidder at the bidding price and supplemented by the project. This programme is co-financed by the public, the Malaysia Energy Commission and the PV industry. It is expected that PV suppliers will finally offer BIPV system prices equivalent to those in Europe and Japan. Nowadays, the cost of a 5-kWp BIPV turn-key roof-top system in Malaysia is about RM 27,000/kWp; thus, a 5 kWp BIPV system is estimated to cost RM 135,000. This will facilitate the creation of a sustainable BIPV market upon the completion of the programme. The system will produce approximately 6,000 kWh of energy per annum [10].

SOLAR HEATING SYSTEMS IN MALAYSIA

Ali et al. [11] noted Malaysia's suitability for harnessing solar energy and developing solar technology due to the availability of sunshine. Domestic hot water systems are very popular in Malaysian households, but because of the high initial cost of solar water-heating systems, many are still opting for electric water heaters, which are affordable and easy to install. However, recent figures show that the annual cost of an electric water heater is greater than the cost of a solar water-heating system in a longer term, after 4-6 years. With Malaysia's increasing population, solar water-heating systems will have long-term economic benefits; they are environmentally friendly and play a role in reducing the country's increasing dependency on foreign oil.

The development of a solar-assisted heat pump (SAHP) for hot water supply was described by Hamllaoui et al. [12] and Chowa et al. [13]. The SAHP system [13] (as shown in Figure 2) adopts the vapour compression-refrigeration cycle. The working fluid (refrigerant) passes through the compressor, condenser, capillary tube, evaporator and solar collector in turn, with corresponding thermodynamic-state points denoted 1-5 respectively. The condenser is basically a water storage tank with a submerged heat transfer coil for refrigerant flow from state point 2 to 3. The liquid refrigerant vapourises at the capillary tube to state 4 and evaporates at the evaporator-collector, where point 5 indicates the saturated vapour condition. The superheated refrigerant vapour at state 1 enters the reciprocating compressor and leaves at a higher temperature and pressure located at point 2 [13]. It was found that it is possible to heat water to 55°C at night with the energy available from the ambient surroundings. Figure 3 compares the annual costs of electric water heaters and solar water heating systems [11].

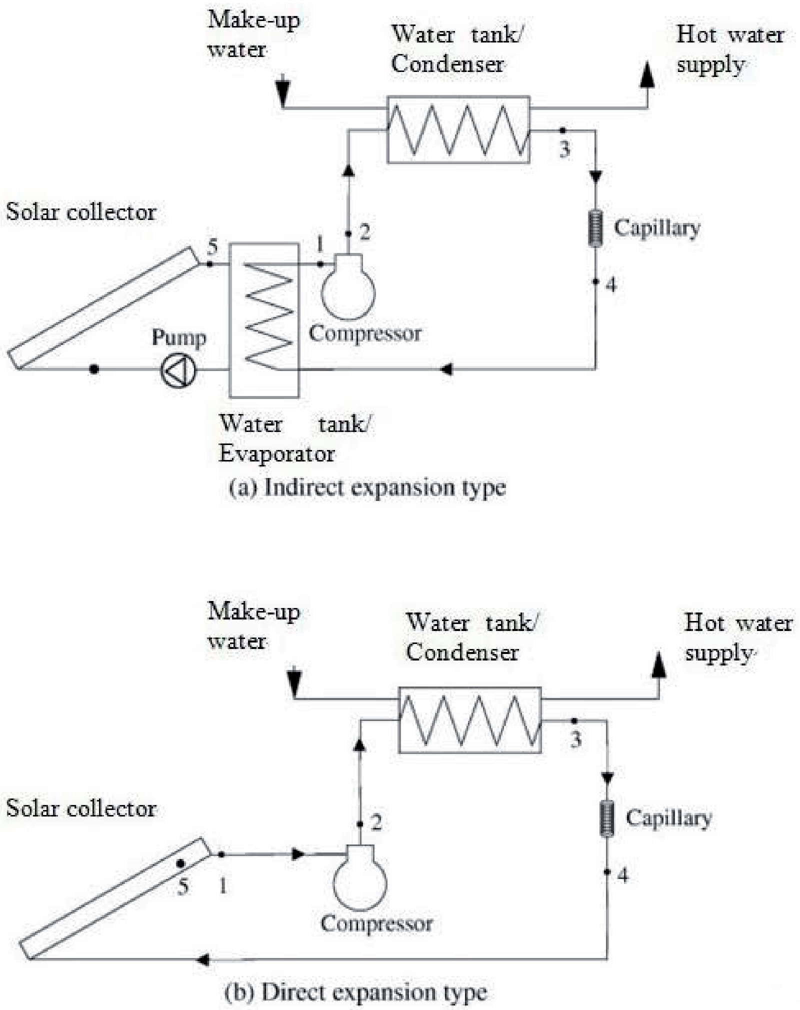


Figure 2. Diagram of SAHP [13]

PASSIVE SOLAR DESIGN IN MALAYSIA

Beggs [14] described a solar-assisted air-conditioning system, a 'mixed-mode' method of addressing the high energy consumption in Malaysia. This system used the heat from solar radiation to drive a thermally-driven chiller such as an absorption chiller, which produced chilled water for indoor cooling purposes. According to MS 1525 [15], a Malaysian standard for green building index, designing with an emphasis on natural daylight should be considered at the preliminary design stage. Conventional and innovative daylight systems that collect, transport and distribute

Figure 3. Comparison of annual costs of electric water heater and solar water heating system [11]. EWH stands for electric water heater and 'M80VTHE' is a common brand of solar water heater whose capacity is 0.285 m^3 and 0.340 m^3 respectively.

light deep into buildings to reduce the need for artificial lighting are recommended. The low energy office (LEO) building in Putrajaya, built as a demonstration project by the Ministry of Energy, Water and Communication, was first occupied in 2004. The building's energy management has been practised since then and the building energy index was aimed at a maximum of 100 kWh /m^2/year. The building energy index in 2005 was 114 kWh/m^2 but it decreased to 104 kWh/m^2 in 2006. An energy audit was done on Block E6 (Ministry of Health) and Block B6 (Economic Planning Unit) of the Prime Minister's Department building in Putrajaya. The audit showed that the energy index of the LEO building was lower than in conventional buildings [16].

For the LEO building, the shade for various facades was optimised using sun charts for geometrical verification and later using ENERGY-10 (a computer application for energy management) for overall optimisation of the design including the effect of daylight on the overall energy balance. The 'punched-hole window' design incorporates a shade element-cum-'light shelf' in the middle of the window. This light shelf is painted with a light-diffusing colour on the upper surface in order to transfer diffuse light into the building through the upper part of the window. The light shelf must have the same depth as the rest of the 'punched-hole' window shade. The punched-window method is used for external windows on Levels 2 and 3, and a curtain wall with exterior shading is used on Levels 4 to 6.

As shown in Figure 4, shading of the atrium is provided by translucent canvas; this is adjusted automatically to reduce sunlight penetration into the atrium and avoid glare in the adjacent office areas. The canvas has a high shading coefficient. A solar sensor which senses the intensity of solar radiation on a horizontal surface is placed above the canvas to control shading. A manual override is provided. The shading and punched-hole windows for the various facades are summarised in Table 1.

Figure 4. Shading device for atrium

Table 1. Shading and punched-hole windows for various facades [17]

Orientation	Shading type	Depth of shading	Shading coefficient	Note
NE	Punched-hole windows	650 mm	0.7 – 0.8	For a 2400-mm-high window with light shelf in the middle
	Louvres / curtain wall	370 mm	0.7	For a distance between louvres of 1000 mm
SE	Punched-hole windows	1000 mm	0.59 – 0.8	For a 2400-mm-high window with light shelf in the middle
	Louvres / curtain wall	820 mm	0.59 – 0.8	For a distance between louvres of 1000 mm
South	Punched-hole windows	650 mm	0.59 - 0.8	For a 2400-mm-high window with light shelf in the middle
	Louvres / curtain wall	370 mm	0.59	For a distance between louvres of 1000 mm
North	Punched-hole windows	1000 mm	0.7 – 0.8	For a 2400-mm-high window with light shelf in the middle
	Louvres / curtain wall	370 mm	0.8	For a distance between louvres of 1000 mm
NW	Punched-hole windows	1000 mm	0.7 -0.8	For a 2400-mm-high window with light shelf in the middle
	Louvres / curtain wall	820 mm	0.7	For a distance between louvres of 1000 mm

To maximise daylight, space in the LEO building is divided into three zones: the daylight zone, no-daylight zone and semi-daylight zone, which are used for different purposes. The office rooms in the daylight zone make good use of daylight and artificial lighting is not needed most of the time. Glare is minimised by the light shelves, especially in the south- and north-orientated offices. The building is deep and artificial lighting is needed in the inner zones where most of the staff work. The atrium provides a certain amount of daylight but not enough to avoid artificial lighting. The light is on constantly during working hours but a lighting control system reduces the electricity demand considerably. The energy demand for lighting is more than double that of the estimated load. On the east facade the morning sun shines directly into the offices and to avoid discomfort blinds are used. Later in the day the direct sun does not enter the offices but the blinds tend not to be removed and the daylight is not used to advantage. Utilisation of daylight is excellent in the offices along the south and north facades and poor in the inner zones of the building.

According to MS1525 [15], the average daylight factors may be obtained by using nomograms, computer simulation or architectural modelling of a building design. Scaled models or computer simulations of buildings with more than 4000 m² of air-conditioned space are used to model daylight performance. Daylight factors were measured in several rooms in the LEO building. The rooms were selected to represent different orientations of the building façade. It was found that in general, without the use of blinds, a daylight factor of more than 2% is available for up to 3-m depth from the perimeter of the glazing, while a daylight factor of 1% is available for up to 6-m depth in the office space of all orientations. Figure 5 provides a view of the lighting effect with blinds.

Figure 5. Lighting effect with blinds

The daylight lux level measurement of various rooms indicates that daylight distribution is even for the first 2 m of depth, gradually decreasing beyond that. The change in the daylight factor of 1-2% between 3-6 m shows that the daylight is well distributed and provides an even daylight environment for rooms less than 7 m deep. The light shelf seems to produce the necessary even light distribution by reflecting some of the light further into the room and reducing the light nearer the glazing. It is interesting to note that a daylight factor of 1% seems adequate for office work here. This may indicate that a high level of daylight is available outside or that people are more willing to accept a lower lighting level when they can see that the outside light level is also low. Further analysis of the available daylight outside the hours of 8 a.m. - 5 p.m. is recommended for further academic research. A study of the occupants' preferences and tolerance of light levels in space lit by natural light where they can see the outside daylight at the same time is also proposed.

Daylight was measured on the afternoon of 14 November 2005 and it was found that blinds were in use in approximately 50% of the rooms, necessitating the use of artificial light. Most of the rooms facing east and some of the rooms facing north and south had the blinds in use. The east facade received direct sunshine in the morning, creating glare and causing the occupants to use the blinds. However, once the blinds were drawn they were rarely opened again, even when the sun had moved away and would no longer cause glare. Some of the blinds in use in the north- and south-facing rooms were to prevent glare on computer screens to the occupants sitting with the window behind them. This practice warrants comprehensive research to gain an understanding of the occupants' behaviour in relation to blind use, visual comfort and energy efficiency.

BIPV TECHNOLOGY IN MALAYSIA

In 2005, the Malaysian government launched its first collaborative effort in using renewable energy technology, specifically PV technology, in its buildings in the Malaysian BIPV project. Unlike in the UK, the BIPV sytem used in Malaysia is grid-connected. Subsequently, in 2006 the government launched the SURIA1000 project focusing on residential BIPV systems. In the 9th Malaysian Plan, the government made a substantial allocation for solar technology application, especially in the east Malaysia region [18]. Figure 6 presents a diagram of a basic grid-connected PV system and Figure 7 illustrates the grid-connected BIPV system concept.

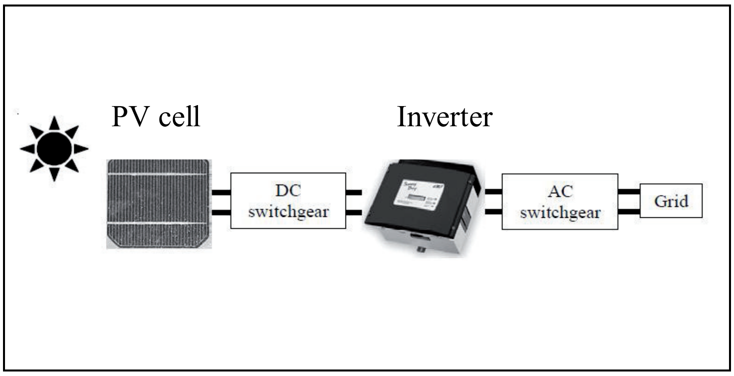


Figure 6. Diagram of basic grid-connected PV system [18]

Sunlight is converted by the PV module into direct current, which is fed into an inverter that converts it into alternating current transferable to the grid. The system responds by importing energy from the grid when energy from the PV system is insufficient, or exporting excess energy to the grid. Apart from government buildings such as the Centre for Environment, Technology and Development and Green Energy Office (GEO) of Pusat Tenaga Malaysia (PTM) as well as Monash University (Sunway), residential areas such as those in Cheras, Semenyih, Bukit Sebukor (Malacca), Setia Eco Park, Putrajaya and Bangsar have also incorporated PV modules into their buildings under the MBIPV project.

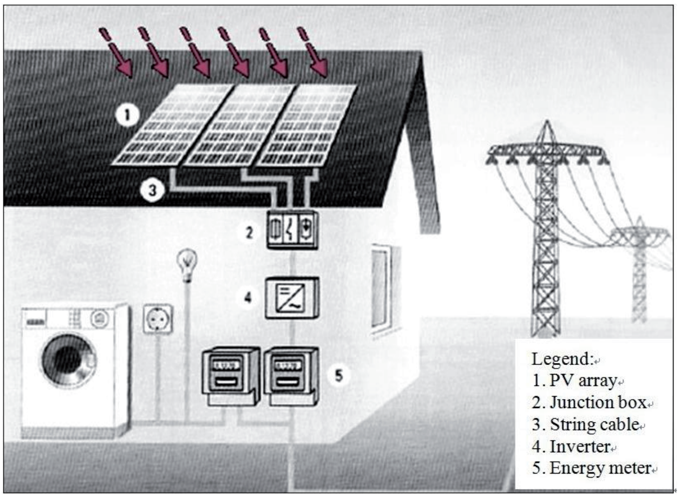


Figure 7. Grid-connected BIPV system concept [18]

The PTM GEO building was developed following the success of the LEO building in Putrajaya. Table 2 summarises some of the project data for the GEO building. Its demonstrates that the Zero Energy Building concept can be realised using technologies already available today and taking advantage of a climate where solar radiation and daylight can be used as the energy sources, at the same time reducing fossil fuel consumption. This building promotes advanced energy technologies for use in buildings in Malaysia, technologies that will ultimately make buildings self-sustaining in energy supply.

Pre-programmed into the building's 'DNA' are energy-efficient features and the BIPV system, which make up the backbone of this self-sufficient, fully sustainable landmark. The PTM's GEO building does not use fossil fuels, driving home the point that an office building need not consume electricity derived from this source. Instead, all the electricity needed by the building is generated by its own solar BIPV systems. Four different systems utilising four different technologies have been installed, as shown in Figure 8. The first and biggest component features a 47.28-kWp polycrystalline BIPV system on the main roof; the second component is a 6.08-kWp amorphous silicon BIPV system incorporated into the second main roof; the third system stored in the atrium of the building highlights the use of a 11.64-kWp monocrystalline glass-glass BIPV system; and lastly, the car park roof is fitted with a 27-kWp monocrystalline BIPV system.

Table 2. Summary of project data for the GEO building [19]

Project name:	Malaysia Energy Centre: Green Energy Office
Location:	Section 9, Bandar Baru Bangi, Bangi, Selangor
Completion:	July 2007
Site area:	2 hectares
Gross floor area:	4,000 m^2
Number of rooms:	41 rooms
Building height:	47.50 metres
Client/Owner:	Malaysia Energy Centre (Pusat Tenaga Malaysia)

Figure 8. Four types of solar panel and their locations on the GEO building [20]

The solar BIPV systems are all linked to grid-connected inverters that convert DC produced into AC. The amount of electricity generated is recorded by a meter. In this case, no battery is installed as the generated solar electricity is consumed directly and the net surplus sold to Tenaga Nasional Berhad (TNB) on a net meter basis. At a total BIPV capacity of 92 kWp, the anticipated target for electricity generated annually by solar BIPV systems stands at 102 MWh [17]. To date, the BIPV systems have produced an estimated 103 MWh/year on average, based on actual output over three months. Buildings that are not energy efficient would need more than the GEO's 92 kWp because the super energy efficient features of the GEO reduce the energy consumption of the building. The total payback time for the whole GEO system should be less than 22 years [20], based on the current subsidised electricity tariff and imported technologies. It is acknowledged that future electricity costs will increase while the cost of energy efficient features and solar technologies will decrease. However, it is important to bear in mind that there is no payback price for the environment.

ENERGY LEGISLATION AND POLICY IN MALAYSIA

Malaysia is a strong supporter of the ASEAN countries' renewable energy development programme, implementing a Five Fuel Strategy under the 8th Malaysian Plan, and regarding renewable energy as the fifth fuel. However, Malaysia is still far from achieving the desired target.

The Malaysia Energy Centre was launched with the aim of playing a major role in developing energy efficiency and renewable energy. Its message was reinforced through its GEO (formerly known as the zero-energy office) and its building was certified as Malaysia's first completed green-rated office building. It was inspired by another energy-efficient building, the LEO building, which is now the Ministry of Energy, Green Technology and Water. Malaysia has set the standards for energy efficiency through these remarkable buildings. The GEO building incorporates solar technology such as BIPV panels, at the same time exploiting solar radiance in the building design for cooling and air convection. The LEO also uses PV systems, although they operate in parallel with the grid supply. The Malaysian government has introduced codes of practice for energy efficiency, hoping to encourage more ‘green’ buildings in the country. These are MS 1525:2001[15], specifying the minimum criteria for being a low energy building, and MS 1837:2005 [21] which is used as guidelines for BIPV-system projects.

Recently, the prime minister of Malaysia, Datuk Seri Mohd Najib, launched the Green Technology Policy 2009, stressing energy efficiency and green building design. Some of the approaches taken were setting up a green technology agency, promoting foreign investments, allocating incentives for students pursuing green technology, and provision for R&D.

Malaysia’s development is tied to three key national policy frameworks: new economic policy (NEP), 1971-1990; national development policy (NDP), 1991-2000, and national vision policy (NVP), 2001-2010. The framework for developing energy is outlined in the national energy policy 1979, national depletion policy 1980 and fuel diversification policy (four-fuel diversification policy 1981 and renewable energy as the fifth-fuel policy 2000). The four-fuel diversification policy 1981 identifies the country’s energy mix as oil, natural gas, coal and hydro power. Due to increasing oil prices and environmental degradation, in 2001 the government of Malaysia introduced the fifth- fuel policy, adding renewable resources to the energy mix with emphasis on sustainability and efficiency. Since then, the government has concentrated on energy efficiency in industrial and commercial sectors as well as residential sectors, and on the utilisation of renewable energy by promoting new renewable energy resources, e.g. biofuel, landfill gas, mini-hydro power and solar energy, and sustainable development of all energy resources [22]. Table 3 traces the history of various energy policies/acts in Malaysia's national energy development.

To further show its commitment to promoting low-carbon technologies and ensuring sustainable development while conserving the natural environment and resources, the Malaysian government launched its national green-technology policy (NGTP) on 24 July 2009 [23]. Green technology refers to the development and application of products, equipment and systems used to conserve the natural environment and resources, minimising the negative impact of human activities and satisfying the following criteria: (i) green technology minimises environmental degradation; (ii) it has zero or low greenhouse-gas emissions; (iii) it is safe to use and promotes a healthy environment for all forms of life; (iv) it conserves energy and natural resources; and (v) it promotes the use of renewable resources [24]. The NGTP has five strategic thrusts as follows: (i) strengthen the institutional frameworks; (ii) provide an environment conducive to green technology development; (iii) intensify human capital development in green technology; (iv) intensify green technology research and innovations; and (v) promote public awareness. The aim of the NGTP is to foster progress and make improvements in major sectors such as energy, buildings, water and waste management, and transportation as detailed in Table 4.

Table 3. Malaysia's policies/acts on national energy development [22]

Policy/act	**Key emphasis**
National Petroleum Policy (1975)	Introduced to ensure optimal use of petroleum resources, regulation of ownership, management and operation, and economic, social and environmental safeguards in the exploitation of petroleum due to fast growing petroleum industry in Malaysia.
National Energy Policy (1979)	Formulated with broad guidelines on long-term energy objectives and strategies to ensure efficient, secure and environmentally sustainable supplies of energy. Three main objectives: 1. Supply objective: To ensure provision of adequate, secure and cost-effective energy supplies through developing indigenous energy resources, both non-renewable and renewable, using the least cost options and diversification of supply sources both from within and outside the country. 2. Utilisation objective: To promote efficient utilisation of energy and to discourage wasteful and non-productive patterns of energy consumption. 3. Environmental objective: To minimise negative impacts of energy production, transportation, conversion, utilisation and consumption on the environment
National Depletion Policy (1980)	Introduced to safeguard against over-exploitation of oil and gas reserves. Thus, it is a production control policy.
Four-Fuel Diversification Policy (1981)	Fuel diversification is designed to avoid over-dependence on oil as main energy supply and aimed at placing increased emphasis on gas, hydro power and coal in the energy mix.
Electricity Supply Act (1990)	Regulates licensing of electricity generation, transmission and distribution.
Gas Supply Act (1993)	Regulates licensing of gas supply to consumers, pricing, installations and appliances as well as safety aspects.
Fifth-Fuel Policy (2000)	Introduced in recognition of the potential of biomass, biogas, municipal waste, solar and mini-hydro power as potential renewable energy sources for electricity generation.
Energy Commission Act (2001)	The Energy Commission was established to provide technical and performance regulation for electricity and piped gas supply industries, and to advise the government on matters related to electricity and piped gas supply including energy efficiency and renewable energy issues. The Electricity Supply Act 1990 and Gas Supply Act 1993 have both been amended to allow the Energy Commission to take over these responsibilities.
National Biofuel Policy (2006)	Supports the five-fuel diversification policy. Aimed at reducing the country's dependence on depleting fossil fuels and promoting demand for palm oil. Five key thrusts: transport, industry, technology, exports and cleaner environment. Highlights: 1. Producing biodiesel fuel blend of 5% processed palm oil with 95% petroleum diesel; 2. Encouraging the use of biofuel by giving incentives for providing biodiesel pumps at fuelling stations; 3. Establishing industrial standards for biodiesel quality under standards and industrial research institute of Malaysia (SIRIM); 4. Setting up of palm oil biodiesel plants.

The prime minister of Malaysia, Datuk Seri Najib Razak, in his keynote speech on 24 July 2009 [25] during the launching of the NGTP, stated that between 2009-2030 global primary energy consumption is expected to rise by 1.6% annually. Malaysia's electricity demand is forecasted to reach 18,947 MW by 2020 and 23,092 MW by 2030, a 35% increase from 14,007 MW in 2008. Currently, Malaysian electricity capacity through renewable energy stands at 50 MW and it is

Table 4. Significant progress and major improvements aimed by the NGTP [24]

Sector	**Progress and major improvements**
Energy supply	Application of green technology in power generation and energy-supply management, including co-generation of power by industrial and commercial sectors
Energy utilisation	Application of green technology in all energy utilising sectors and demand-side management programmes
Buildings	Adoption of green technology in the construction, management, maintenance and demolition of buildings
Water and waste management	Adoption of green technology in the management and utilisation of water resources, wastewater treatment, solid waste and sanitary landfill
Transportation	Incorporation of green technology, particularly biofuels and public road transport, into transport infrastructure and vehicles

expected to reach about 2000 MW by 2020. Thus, the Malaysian government will institute a renewable energy law in 2011, including a policy focusing on a feed-in tariff (FiT) mechanism. Different forms of renewable energy are given their share in the FiT mechanism as shown in Table 5. The proposed duration for the FiT ranges between 16-21 years with an annual digression of up to 6%. Less focus seems to be given to mini-hydro power and more focus to solar energy. The policy trends towards promotion of renewable energy have shown that Malaysia is aware of the importance of renewable energy and energy efficiency in its energy development. Based on the energy trends, the national economic developments are similarly aligned in the Malaysia Plans. Table 6 shows the key emphases in energy developments from the 7^{th}-10^{th} Malaysia Plans.

Table 5. Proposed FiT rates [26]

Renewable energy (RE)	**Duration (year)**	**Tariff (RM/kWh)/(USD/kWh)**	**Annual digression**	**Displaced electricity cost (RM/kWh)**
Wind	21	0.23–0.35/0.071–0.108	1%	0.22
Solar PV	21	1.25–1.75/0.386–0.541	6%	0.35
Solid waste and sewage gas	21	0.3–0.46/0.093–0.142	1.5%	0.22
Biomass	16	0.24–0.35/0.074–0.108	0.2%	0.22
Biogas	16	0.28–0.35/0.087–0.108	0.2%	0.22
Geothermal	21	0.28–0.46/0.087–0.142	1%	0.22
Mini-hydro	21	0.23–0.24/0.071–0.074	0%	0.22

Table 6. Malaysia's key emphases for energy development [26]

Malaysia Plan	**Key emphases**
Seventh Malaysia Plan (1996–2000)	- Emphasis on sustainable development to compensate for depletable resources, and on diversification of energy sources. - Ensuring adequacy of generating capacity as well as expanding and upgrading transmission and distribution infrastructure. - Encouraging the use of new and alternative energy sources as well as efficient utilisation of energy.
Eighth Malaysia Plan (2001–2005)	- Emphasis on sustainable development of energy resources, both depletable and renewable. The energy mix includes five fuels: oil, gas, coal, hydro and renewable energy (RE). - Intensifying efforts on ensuring adequacy, quality and security of energy supply. - Greater emphasis on energy efficiency (EE): encouraging efficient utilisation of gas and RE as well as providing adequate electricity generating capacity. - Supports development of industries in production of energy-related products and services. - Highlights promotion of RE and EE. - Incentives for EE. - Incentives for use of RE resources. - Incentives to maintain quality of power supply.
Ninth Malaysia Plan (2006–2010)	- Emphasis on strengthening initiatives for EE, especially in transport, commercial and industrial sectors, and in government buildings. - Encouraging better utilisation of RE through diversifying fuel sources. - Intensifying efforts to further reduce dependency on petroleum and to integrate alternative fuels. - Incentives to promote RE and EE are further enhanced.
Tenth Malaysia Plan (2011–2015)	Short-term goals vested in NGTP: - Increased public awareness and commitment for adoption and application of green technology through advocacy programmes. - Widespread availability and recognition of green technology in terms of products, appliances, equipment and systems in local market through standards, rating and labelling programmes. - Increased foreign and domestic direct investments in green technology manufacturing and service sectors. - Expansion of local research institutes and institutions of higher learning to accommodate research, development and innovation activities on green technology with the aim of commercialisation through appropriate mechanisms. - New RE act and FiT mechanism to be launched.

CHALLENGES OF SOLAR ENERGY IMPLEMENTATION

Energy efficiency means using less energy to produce the same amount of services or useful output [27, 28]. The rapid growth of energy use has already raised concern over problems of supply worldwide. It is exhausting the energy resources and causing severe environmental impacts such as depletion of the ozone layer, global warming and climate change. In developed countries, especially in Malaysia, the contribution from residential and commercial buildings to energy consumption has steadily increased to between 20-40% of the total energy consumption [29]. Nowadays, energy efficiency of buildings is a prime objective for energy policies at regional, national and international levels due to population growth and increasing pressure on building services to handle the continuing upward trend in energy demand [30,31]. The current energy and socio-economic systems are clearly unsustainable and in generating electricity for commercial and domestic uses will undoubtedly exhaust fossil fuel resources. The lifestyles of the developed nations will have a serious environmental impact and will lead to patterns of consumption with increasing energy needs and consequently to higher carbon dioxide concentrations in the atmosphere [32,33]. Energy efficiency is only one of many qualities valued in a building and some of these, such as the use of large glass surfaces, may even counteract energy efficiency [34]. The rapid growth in energy demand is expected not only in developed countries but also in developing countries as they attempt to reach a higher living standard [35,36]. Efficient use of energy will play an essential role in minimising energy usage and associated emissions released into the atmosphere [37].

One of the most cost-effective measures in reducing emissions of carbon dioxide, believed to be a major cause of global warming, is the improvement of energy efficiency in buildings [38,39]. To comply with measures to save energy and reduce environmental pollution, the Malaysian government has already built the LEO building, whose energy intensity is 114 kWh/m^2, only 50% of that in conventional buildings [40]. However, this model building has very few followers. In encouraging the use of renewable energy, the Malaysian government has included a target for renewable energy use of 5% in its 9th Malaysia Plan to reduce the burden on the atmosphere. In addition, University Science Malaysia's centre for education and training in renewable energy and energy efficiency is playing an active role in increasing awareness of energy efficiency and the use of renewable energy among end users [29]. The ineffective distribution of costs and benefits between actors makes the promotion of energy efficiency in the new construction particularly complicated in Malaysia [41]. For residents, energy costs are a small and well-hidden part of the total rent [42]. In relation to supply strategies, there may be a number of reasons why energy efficiency often falls short; one is the problem of imperfect information, which is typically more pronounced in the market of energy efficiency. The service of energy efficiency may consist of a series of small changes with unclear estimates of costs and benefits. Unlike energy supply issues, which are driven by major actors, energy efficiency lacks influential advocates [43]. Furthermore, householders or landlords who have made the investment in energy efficiency generally require much shorter payback time than that for investments in the extension of supply made by major specialised energy companies [33].

The issue of sustainable development and renewable energy in buildings is frequently addressed, yet only a minority of Malaysia's population are actually putting words into action. Like all new developments, solar technology in Malaysia faces several issues and challenges:

Lack of policy framework. Despite the aims of the fifth-fuel policy and green technology policy, there is a lack of comprehensive detail. These policies merely state their objectives and numerous approaches the government can use to create a greener, sustainable development in the future. MS 1525:2001 is effective, yet it does not ensure that all future constructions will be low energy consumption buildings, while MS 1837:2005 has yet to encourage confidence in the public.

Expensive solar technology and lack of expertise. Despite rapid development in solar technology, it is still in its infancy in Malaysia. Installation of solar panels is very costly even for small projects, and there are only a few specialists or consultants who truly understand and know how to effectively save energy in the long term. Because of the high capital investment required, developers in the private sector do not even consider solar technology in their projects.

Lack of public interest and awareness. Public interest at a national level can be considered as relatively low. Environmental and energy issues are generally ignored because the public take it for granted that the government will propose a solution when the problems arise. With fossil and other non-renewable fuels still available to the public, alternative technologies tend to be overlooked.
Insufficient funding. Despite government incentives, funding is still insufficient for most of the projects in Malaysia to consider energy efficient buildings. There are no specific loans from the financial sector to encourage the development of solar technology, and the private sector is always more profit-conscious in the shorter term rather than the long term.

POSSIBLE SOLUTIONS TO CHALLENGES

A future with limited amounts of non-renewable fuels is drawing nearer each day. All the issues raised should be dealt with comprehensively before it is too late:

Legislation and policy. In order to implement solar technology in buildings, acts and policies regarding incentive subsidies should be passed for all future developments. The legislation should include incentive subsidies for: 1) solar technology such as the BIPV system connected to the grid supply for large-scale developments; 2) thin-film solar panels to be incorporated in residential buildings or small projects; and 3) renewable energy as a preferred component of electricity generation for use in homes and other buildings.
Funding from the Private Sector. Private developers and financial institutions are very influential in the construction industry and can potentially affect the solar technology market. Developers should be instilled with the awareness of environmental issues and the need for energy efficiency in buildings, while self-motivation to use solar technology in their projects should be encouraged. Financial institutions could offer developers loans with special interest rates for developments incorporating solar technology. A 'national solar technology fund' could be set up in a campaign to create more awareness among the public and to encourage people to seek out solar technology for domestic purposes.
Tax relief or rebates. The government plays a key role in encouraging the public to enter the solar technology market by giving tax relief or rebates to those who implement solar energy conversion. Tax relief can appeal to the public and create an interest. Furthermore, rebates can be given on electricity bills if their daily energy generation and consumption are from a renewable energy source.
Skilled and knowledgeable manpower. Undoubtedly, skilled manpower or labour in this industry is crucial and will play an important role in the future progress of the construction projects. Malaysia seems to be lacking in specific programmes; most educational institutions regard solar technology as a part of a wider subject, and not as a programme in its own right. The scope of solar technology is wide and there is much to explore and develop, but the provision of skilled manpower is not keeping pace. One way to overcome this is to propose more programmes on the subject at a higher educational level. Technical or vocational schools and comprehensive diploma programmes in universities can produce the required expertise in this field. Workshops and courses can be organised on a weekly basis for professionals such as building surveyors, engineers and architects.

CONCLUSIONS

Solar technology as an energy efficient approach in buildings is generally considered to be beneficial and would result in a cleaner environment in the long term. Stricter legislation is vital. Rules and standards will ensure conformity in future developments. Furthermore, the cost of solar technology should be controlled and be in proportion to the project size. Funding from private organisations can be a huge influence in the country's advancement of solar technology and its use as well as the progress towards sustainable development. Tax relief and rebates will attract more public awareness and instil interest in energy-saving techniques in daily life. An opportunity exists at the educational level for knowledgeable and skilled manpower to be produced on a large scale through research programmes and other educational activities. Malaysia has so much potential for solar technology, especially in the construction industry. With environmental issues and a worldwide energy crisis looming, solar technology is surely one of the ways forward to a greener and more sustainable development in the future. Malaysia faces several issues and challenges in accepting new developments, especially solar technology, such as the lack of a policy framework, lack of expertise, lack of public interest and awareness, and insufficient funding. Appropriate policies to encourage the implementation of solar energy are recommended for future study.

REFERENCES

1 M. Z .A. A. Kadir and Y. A. Rafeeu, "A review on factors for maximizing solar fraction under wet climate environment in Malaysia", *Renew. Sust. Ener. Rev.*, **2010**, *14*, 2243-2248.

2 A. O. Adelaja, O. Damisa, S. A. Oke, A. B. Ayoola and A. O. Ayeyemitan, "A survey on the energy consumption and demand in a tertiary institution", *Maejo Int. J. Sci. Technol.*, **2008**, *2*, 331-344.

3 A. W. Azhari, K. Sopian, A. Zaharim and M. Al-Ghoul, "A new approach for predicting solar radiation in tropical environment using satellite images - Case study of Malaysia", *WSEAS Trans. Environ. Dev.*, **2008**, *4*, 373-378.

4 J. Goldemberg, "The case for renewable energies", in "Renewable Energy: A Global Review of Technologies, Policies, and Markets" (Ed. D. Abmann, U. Laumanns and D. Uh), Earthscan, London, **2006**.

5 A.R. Mohamed, and K.T. Lee, "Energy for sustainable development in Malaysia: energy policy and alternative energy", *Energ. Policy,* **2006**, *34(15)*, 2388-2397.

6 S. Durdic, S. Stojkovic and D. Sabic, "Nature conservation in urban conditions: A case study from Belgrade, Serbia", *Maejo Int. J. Sci. Technol.*, **2011**, *5*, 129-145.

7 A. F. Jamaludin, "Energy mix and alternatives energy for sustainable development in Malaysia", Proceedings of 9th International Students Summit on Food, Agriculture and Environment in the New Century, **2009**, Tokyo, Japan, pp.1-9.

8 Ministry of Energy Communication and Multimedia Malaysia, "Global environment facility project brief", United Nations Development Program, **2010**, http://www.mbipv.net.my/content.asp?zoneid=1&categoryid=10 (Accessed: 5 Dec. 2010).

9 Malaysia Energy Centre, "Malaysia building integrated photovoltaic (MBIPV) technology application project", **2010**, http://eib.ptm.org.my/ (Accessed: 10 Dec. 2010).

10 Malaysia Energy Centre, "Suria 1000 Programme", **2010,** http://www.mbipv.net.my/content.asp?higherID=5&zoneid=4&categoryid=6&hghid=5 (Accessed: 1 Dec. 2010).

11 B. Ali, K. Sopian, M. Al-Ghoul, M. Y. Othman, A. Zaharim and A. M. Razali, "Economics of domestic solar hot water heating systems in Malaysia", *Eur. J. Sci. Res.*, **2009,** *26*, 13-21.

12 A. Hamlaoui, K. S. Ong, T. C. Fah and H. H. Masjuki, "Development of a direct expansion solar assisted heat pump for hot water supply", Proceedings of World Renewable Energy Congress, **1999**, Kuala Lumpur, Malaysia, pp.139-141.

13 T. T. Chow, G. Pei, K. F. Fong, Z. Lin, A. L. S. Chan and M. He, "Modeling and application of direct-expansion solar-assisted heat pump for water heating in subtropical Hong Kong", *Appl. Ener.*, **2010,** *87*, 643-649.

14 C. Beggs, "Energy: Management, Supply and Conservation", Butterworth-Heinemann, Oxford, **2002,** pp.243-266

15 MS1525: Code of practice on energy efficiency and use of renewable energy for non-residential buildings, Standard and Industrial Research Institute of Malaysia Berhad, Malaysia, **2001**.

16 Z. A. Azni, "Integrating sustainable energy in buildings: A case study in Malaysia" Proceedings of Foreningen af Udviklingsforskere (FAU) Conference, **2008**, Copenhagen, Denmark, pp.78-91.

17 P. E. Kristensen, "Towards zero carbon buildings in South-East Asia", Proceedings of IEA Sustainable Buildings Network Conference, **2010,** Paris, France, pp.1-13.

18 S. Shaari, A. M. Omar, H. C. M. Haris, Z. M. Zain, H. Zainuddin, S. I. Sulaiman, K. S. Muhammad, M. Mustaffa, Z. M. Darus, W. F. Abbas, R. A. H. A. Kadir, L. Rimon, P. Mazlan, A. R. Ghazali and R. A. Rahman, "Applying renewable energy technology in Malaysia: Case study for building integrated photovoltaic", **2008,** http://www.scribd.com/doc/13455539/ applying-renewable-energy-technology-in-malaysia-case-study-for-building-integrated-photovoltaics/ (Accessed: 13 Oct. 2010).

19 H. Ibrahim, "Energy efficient buildings", *Ener. Smart*, **2005**, *1*, 1-9.

20 Y. Yong, "Pusat Tenaga Malaysia's Zero Energy Building", **2010,** http://www.futurarc.com/previous_edition/zeroenergy.cfm (Accessed: 9 Feb. 2010).

21 MS1837: Installation of grid connected photovoltaic (PV) system, Standard and Industrial Research Institute of Malaysia Berhad, Malaysia, **2005**.

22 S. C. Chua and T. H. Oh, "Review on Malaysia's national energy developments: Key policies, agencies, programmes and international involvements", *Renew. Sust. Ener. Rev.*, **2010**, *14*, 2916-2925.

23 Bernama, "National green technology policy launched", News Straits Times Press, July 24, **2009**, Kuala Lumpur, Malaysia.

24 Green Purchasing Network Malaysia, "Definition of green technology by KETTHA", Ministry of Energy Technology and Water (Malaysia), **2009**, http://www.gpnm.org (Accessed: 15 Nov. 2009).

25 N. Razak, "Launch of the national green technology policy and official opening of the Malaysia Energy Centre New Building", **2010**, http://www.pmo.gov.my/?menu=speech&page=1908&news_id=153&speech_cat=2 (Accessed: 3 Oct. 2010).

26 A. H. Haris, "Feed-in-tariff (FiT): Driving forward green technologies and deployments", Pusat Tenaga Malaysia, **2009**, http://www.mbipv.net.my/dload/NPVC%202009/Ir.%20Ahmad%20Hadri%20Haris.pdf (Accessed: 30 May 2010).

27 C. Karkanias, S. N. Boemi, A. M. Papadopoulos, T. D. Tsoutsos and A. Karagiannidis, "Energy efficiency in the Hellenic building sector: An assessment of the restrictions and perspectives of the market", *Ener. Policy,* **2010**, *38*, 2776-2784.

28 W. Chung, Y. V. Hui and Y.M. Lam, "Benchmarking the energy efficiency of commercial buildings", *Appl. Ener.*, **2006,** *83*, 1-14.

29 R. Saidur, "Energy consumption, energy savings, and emission analysis in Malaysian office buildings", *Ener. Policy,* **2009**, *37*, 4104-4113.

30 O. M. Al-Rabghi and M. M. Akyurt, "A survey of energy efficient strategies for effective air conditioning", *Ener. Convers. Manage.*, **2004**, *45*, 1643-1654.

31 J. O. Jaber, M.S. Mohsen, A. Al-Sarkhi and B. Akash, "Energy analysis of Jordan's commercial sector," *Ener. Policy,* **2003**, *31*, 887-894.

32 L. Perez-Lombard, J. Ortiz and C. Pout, "A review on buildings energy consumption information", *Ener. Build.*, **2008**, *40*, 394-398.

33 A. L. A. Olanrewaju, M. F. Khamidi and A. Idrus, "Quantitative analysis of defects in Malaysian university buildings: Providers' perspective", *J. Retail Leisure Prop.*, **2010**, *9*, 137-149.

34 J. Nassen and J. Holmberg, "Energy efficiency - A forgotten goal in the Swedish building sector", *Ener. Policy,* **2005**, *33*, 1037-1051.

35 N. Yamtraipat, J. Khedari, J. Hirunlabh and J. Kunchornrat, "Assessment of Thailand indoor set-point impact on energy consumption and environment", *Ener. Policy*, **2006**, *34*, 765-770.

36 H. Radhi, "A systematic methodology for optimizing the energy performance of buildings in Bahrain", *Ener. Build.*, **2008**, *40*, 1297-1303.

37 B. Farhanieh and S. Sattari, "Simulation of energy saving in Iranian buildings using integrative modelling for insulation", *Renew. Ener.*, **2006**, *31*, 417-425.

38 L. C. Chow, "A study of sectoral energy consumption in Hong Kong (1984-97) with special emphasis on the household sector", *Ener. Policy*, **2001**, *29*, 1099-1110.

39 Y. Uchiyama, "Present efforts of saving energy and future energy demand/supply in Japan", *Ener. Convers. Manage.*, **2002**, *43*, 1123-1131.

40 B. Abd Malek and S. Lojuntin, "Sustainable Building Policies on Energy Efficiency: Malaysia", United Nations Environment Programme (UNEP) and Building and Construction Authority (BCA), Singapore, **2011**, pp.39-41.

41 Z. A. Azni, A. R. Samirah and S. Sulaiman, "Energy in Buildings: Sustainable Symbiosis", University Publication Center (UPENA), Universiti Teknologi MARA, Shah Alam (Malaysia), **2006**.

42 M. Ryghaug and K. H. Sorensen, "How energy efficiency fails in the building industry", *Ener. Policy*, **2009**, *37*, 984-991.

43 S. N. Kamaruzzaman, S. A. Azlan, A. S. Zulkiflee and E. M. A. Zawawi, "Energy performance of electrical support facilities: The case of adaptive re-used historical buildings in Malaysia", *Int. J. Phys. Sci.*, **2009,** *4*, 752-757.

Maejo Int. J. Sci. Technol. **2012**, *6*(02), 216-223

Maejo International Journal of Science and Technology

ISSN 1905-7873
Available online at www.mijst.mju.ac.th

Communication

A freshwater red alga, *Thorea clavata* Seto et Ratnasababapathy, from Thailand with special reference to sexual reproductive organs

Siripen Traichaiyaporn [1,*], Shigeru Kumano [2] and Boonsom Waraegsiri [3]

[1] Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand
[2] National Institute for Environmental Studies, Tsukuba 305-8506, Japan
[3] Department of Agricultural Extension, Faculty of Agricultural Production, Maejo University, Chiang Mai 50290, Thailand

* Corresponding author, email: tsiripen@yahoo.com; tel: (6653) 892-258 ext. 1104, 1132; (6681) 5954532; fax: (6653) 892-259

Received: 18 August 2011 / Accepted: 19 May 2012 / Published: 28 May 2012

Abstract: *Thorea clavata* is reported for the first time in Thailand. Prior to this discovery it was known only from the type locality in Malaysia. The Thai collection was from the sub-district of Huay Khayeng, Thong Pha-phum district, Kanchanaburi province, south-western Thailand. The thalli were fertile and the spermatangia, carpogonia and carposporangia are described and illustrated.

Keywords: freshwater red algae, *Thorea clavata*, algal sexual reproductive organs

INTRODUCTION

The Taxa of *Thorea* show worldwide distribution but tend to be more common in tropical and subtropical regions as well as warmer waters in temperate regions [1]. From the western Pacific region, *T. gaudichaudii* was first discovered on Guam in the Marianas Islands [2] and subsequently reported from Okinawa in Japan [3] and Gagil-Tamil in the Caroline Islands [4]. *T. okadae* was described from Sendai River, Kagoshima prefecture on the island of Kyusyu in Japan [3]. *T. hispida* (as *T. ramosissima*) was reported from China [3]. *T. clavata* was described from Sungai Gombak in Selangor State and *T. prowsei* from Pahang State in the Malaysian peninsula without any description of the sexual reproductive organs [5]. The taxonomy and distribution of the taxa of Thoreaceae in the Western Pacific region were reviewed by Seto [6]. These new species were included in a review of morphological and phylogenetic features by Kumano et al. [7], who recognised five distinct taxa in the Asia-Pacific region: *T. hispida, T.okadae, T. guadichaudii, T. clavata* and *T. siamensis*. The last taxon was described as a new species [8].

In this paper *T. clavata* is reported for the first time in Thailand with a description of sexual reproductive organs.

MATERIALS AND METHODS

Specimens Examined

(1) Specimen from Sungai Gombak, Selangor State, Malaysia, coll. Ratnasabapathy (Herbarium of National Science Museum, Tsubuka, Japan, TNS-AL-157043, isotype).

(2) Specimen from Huay Khayeng sub-district, Thong Pha-phum district, Kanchabnaburi province, Thailand, coll. Traichaiyaporn et al. 8,11 (Herbaium of Department of Biology, Faculty of Science, Chiang Mai University, CMU).

Methods

Pieces of the herbarium specimens were well moistened with distilled water and removed for examination in a similar manner to fresh specimens from the field. Photomicrographs of specimens were taken by means of an Olympus BU-2 microscope, an adapter Olympus U-PMTVC, a CAMEDIA C0304-ADL digital camera and a TV-monitor.

For each specimen, some taxonomic parameters previously used to distinguish infrageneric taxa were measured. Since some type specimens were preserved and often fragmented, the sizes of gametophytes measured did not show exact values. Vegetative structures, e.g. length of gametophyte, cell number, and size and shape of assimilatory filament, were measured. Characteristics of sexual reproductive organs, e.g. size of spermatangia, carpogonia with trichogyne and asexual one such as carposporangia, were determined.

Using data of such parameters, calculation of average standard deviations was made by an \`Excel\` 2003 for Windows.

TOPOGRAPHY AND ENVIRONMENTS IN THAILAND

Thorea clavata is a new record of a freshwater red alga in Thailand, which is the second locality known for this species. It was found in a stream flowing through a mixed deciduous forest at about 205 m above mean sea level (UTM 0453312, 1617355 (N,S)), in Huay Khayeng sub-district, Thong Pha-phum district, Kanchanaburi province (Figure 1). It attached itself to rocks (Figure 2A) in slow-flowing water. The stream was 2.50-3.50 m wide with medium to high light intensity. Physico-chemical water qualities recorded in November 2003 and January 2004 were as follows: colour crystal clear, depth 0.20-1.00 m, velocity 0.09-0.31 m sec^{-1}, temperature 23.7-24.9°C, pH 7.4-7.5, conductivity 344-347 µS cm^{-1}, hardness 66.3-152.7 mg L^{-1} as $CaCO_3$, nitrate-nitrogen < 0.001-1.200 mg L^{-1}, nitrite-nitrogen < 0.001-0.006 mg L^{-1}, orthophosphate phosphorus 0.09-0. 20 mg L^{-1}, and silica 9.1-9.2 mg L^{-1}.

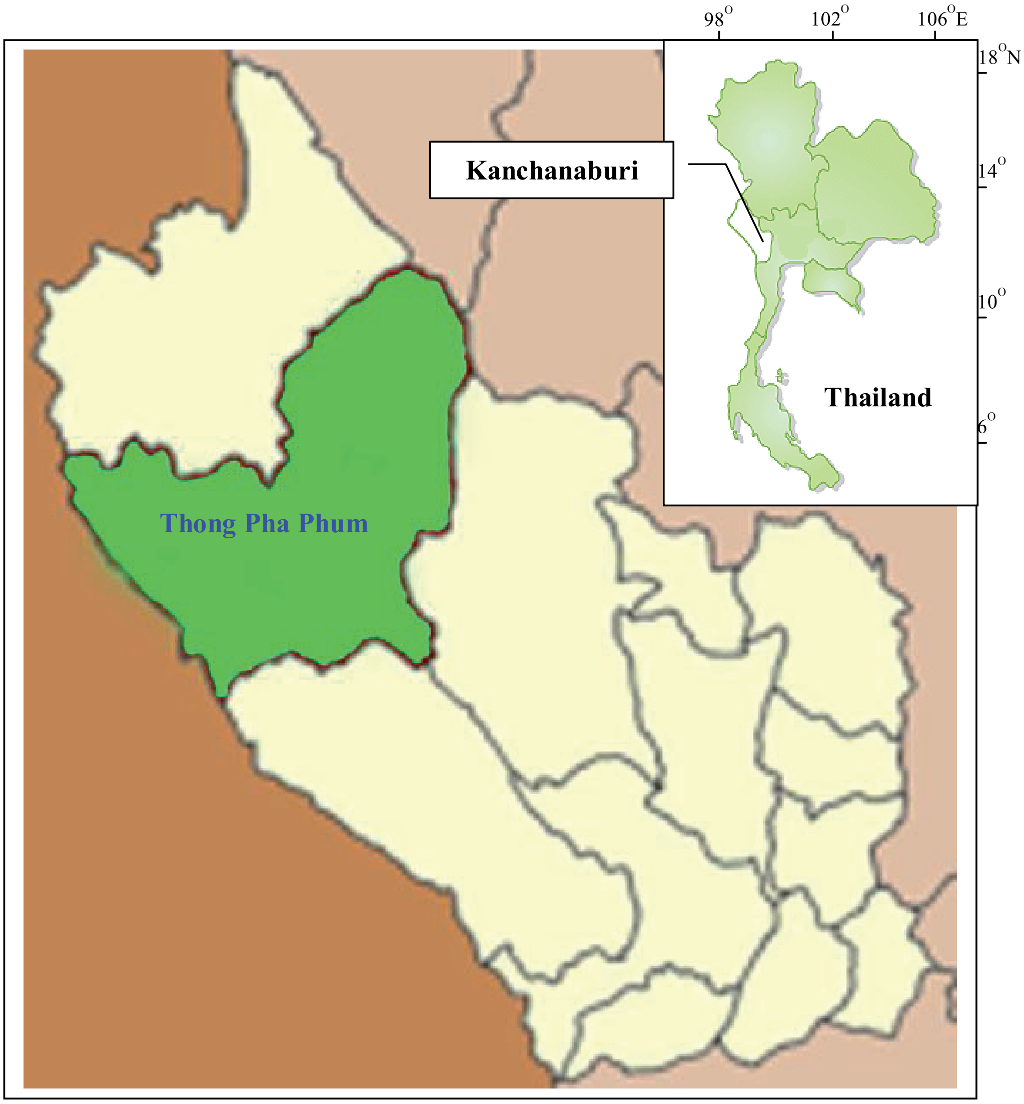


Figure 1. Map of Thailand showing the location of Thong Pha-phum district, Kanchanaburi province

DESCRIPTION OF THE THAI SPECIMEN

Thallus rather slender, tufted, highly mucilaginous, 500-600 µm in diameter, 4-12 cm in length, dark green, abundantly branched, multiaxial, consisting of medullar filaments and cortical assimilatory filaments, attached to substrata with discoid holdfasts. Medullar portion 200-400 µm in diameter (Figures 2B-2C). Assimilatory filaments 203 (170-310) µm in length, consisting of 12.5 (11-14) cells; apical cells clavate with rounded apices (Figure 2D). Distal portion of assimilatory filaments unbranched or sparsely branched, clavate, gradually tapered from apex toward proximal portion (Figure 3A).

Spermatangia terminal in small clusters on short assimilatory branches, elongated ovoid (with diameter/length ratio 1.79), 4.4 (4.3-5.7) µm in diameter and 7.8 (7.2-10) µm in length (Figures 3A-4A).

Carpogonia elongated ovoid (with diameter/length ratio 1.9), 3.7 (2.9-4.3) µm in diameter at the base, 7.2 (7.2) µm in length, trichogyne straight or slightly curved, elongated club-shaped, 1.3 (0.7-4.3) µm in diameter and 95.8 (18-216) µm in length (Figure 4B).

Carposporangia solitary or in clusters terminated on gonimoblast filaments, elongated ovoid, (with diameter/length ratio 1.96), 7.7 (5.7-8.6) µm in diameter, 15.2 (12.9-17.2) µm in length (Figures 3B, 3C).

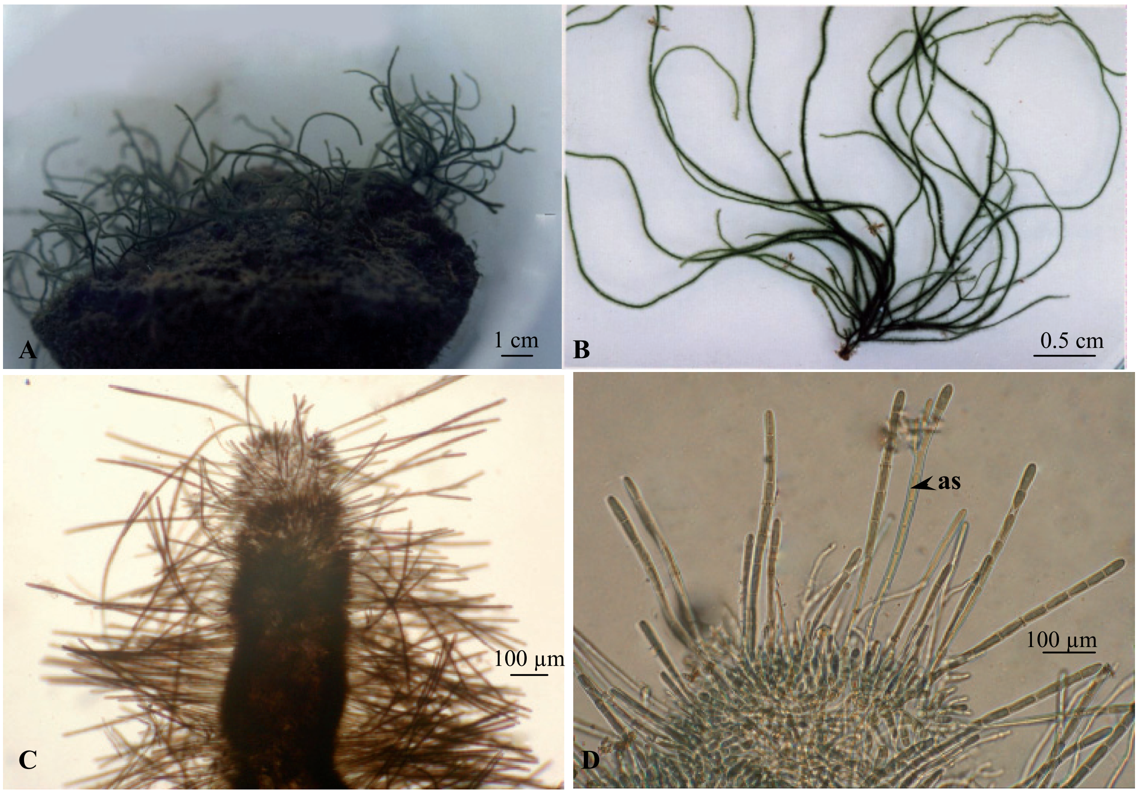


Figure 2. *Thorea clavata* Seto et Ratnasabapathy: (A) plant attaching on rock; (B) whole plant; (C) thallus; (D) assimilaroty filaments (as)

Figure 3. *Thorea clavata* Seto et Ratnasabapathy: (A) assimilatory filaments (as) gradually tapered from apex towards proximal portion; (B) spermatangia (sm) and clusters of carposporangia (cr); (C) clusters of carposporangia (cr)

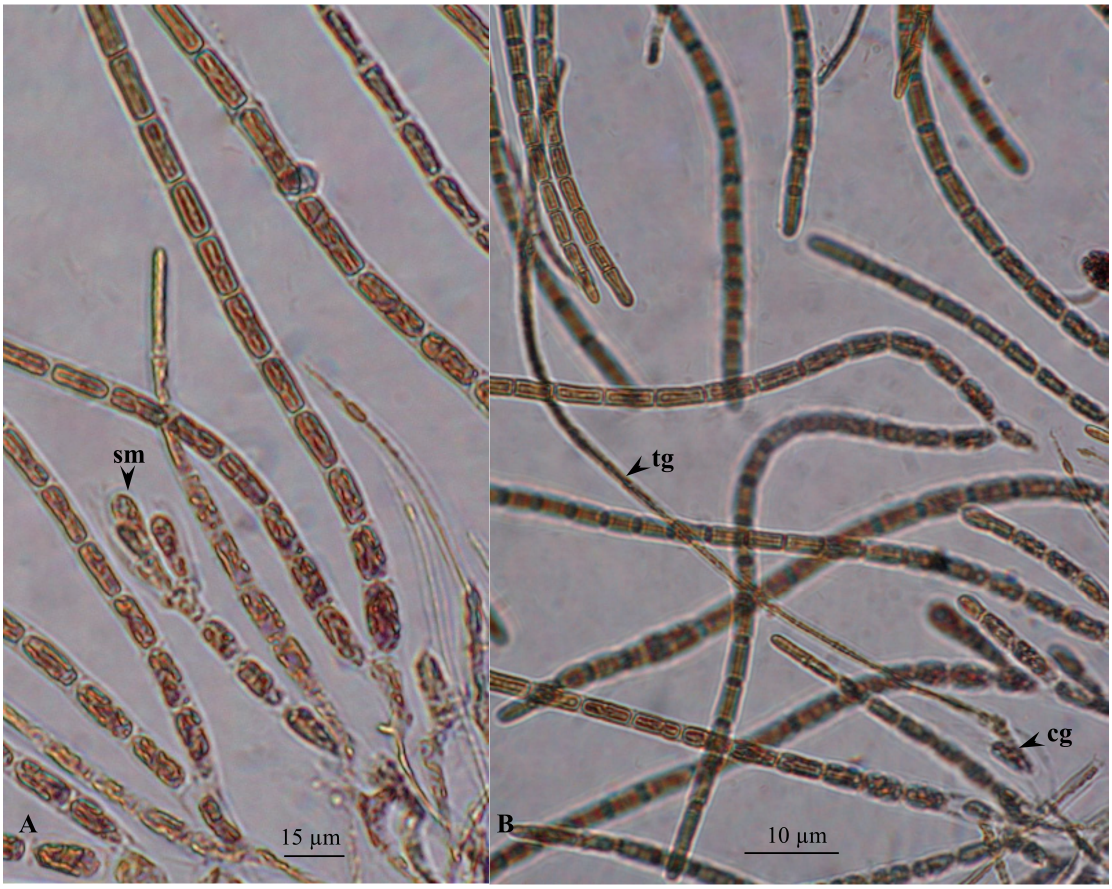


Figure 4. *Thorea clavata* Seto et Ratnasabapathy: (A) spermatangia (sm); (B) carpogonium (cg) with trichogyne (tg)

TAXONOMIC NOTE

The morphological and phylogenetic analyses of taxa of *Thorea* [7] are shown as follows.

Based on shape of assimilatory filaments, two groups of the taxa of genus *Thorea* may be distinguished: *Thorea clavata* with clavate assimilatory filaments and other taxa without clavate assimilatory filaments.

T. clavata was previously reported from Malaysia (type specimen, Ratnasabapathy & Seto 1981) [5].

Phylogenetic analyses based on the *rbc*L and *tuf*A genes show two clades [7]: 1) *T. clavata* from Thailand, *T. violacea* from USA, *T. hispida* from Japan and UK, and *T. okadae* from Japan; and 2) *T. gaudichaudii* from Japan and Philippines, *T. violacea* and *T. riekei* from USA and *T. siamense* from Thailand.

The conclusions are as follows: 1) *T. hispida* and *T. okadae* are sisters in the second clade; 2) the taxa from Japan are regarded as three distinct phylogenetic groups: *T. hispida* group, *T. okadae* phylogenetic group and *T. gaudichaudii* phylogenetic group; 3) *T. clavata* from Thailand is at a basal position in the first clade, closely related to *T. violacea*, *T. hispida* and *T. okadae*; and 4) *T. siamense* from Thailand forms a well-supported subclade as a sister group of *T. gaudichaudii.*

IDENTIFICATION

Rantasabapathy and Seto [5] described *T. clavata* as a new species from Gombak River, Selangor State in Malaysia and mentioned that assimilatory filaments were 130-840 µm in length consisting of 8-40 cells, the distal portion of assimilatory filaments being unbranched or sparsely branched, clavate, gradually tapered from apex towards the proximal portion.

In the present study, the isotype specimen (TNS-AL-157043) gave the following observation: assimilatory filaments 290 (200-400) µm in length consisting of 13.3 (11-16) cells, distal portion of assimilatory filaments unbranched or sparsely branched, clavate, gradually tapered from apex towards proximal portion.

The specimens from Thailand showed the following: assimilatory filaments 203 (170-310) µm in length consisting of 12.5 (11-14) cells, apical cells clavate with rounded apices. Distal portion of assimilatory filaments unbranched or sparsely branched, clavate, gradually tapered from apex towards proximal portion.

Comparing measurements for the three specimens above, the lengths of assimilatory filaments are somewhat different: they are largest for the original description, intermediate for the isotype specimen and smallest for the Thai specimen, with those of the last two being similar to each other. However, based on the common features of the clavata assimilatory filaments, the Thai specimen was identified as *T. clavata*.

HABITAT

In Gombak River, the Malay specimen grew on downstream or upper surface of granitic rocks, 5-40 cm below the surface of clear, unpolluted, fresh, relatively fast or slow stream water flowing through a primary hill country forest with overhanging riverine vegetation. The stream at the site of collection was 2-3 m wide with occasional wide breaks in the vegatation admitting much sunlight. The water temperature at the time of collection (9:30-10:30 am) was 22-22.1°C and the pH was 6.6.

In Kanchanaburi, the Thai specimen was found growing in a small stream, attached on plants and rocks at half-shaded places with medium to high light intensity, in a crystal-clear, shallow small stream 20-40 cm deep, 2.50-3.50 m wide, running through a tropical mixed deciduous forest at an altitude of about 205 m.

TYPE LOCALITY: Gombak River, Selangor, Malaysia.

HOLOTYPE: Herb. Ratnasabapathy, RS 490, Seto, 6/V 1978.

ACKNOWLEDGEMENTS

We wish to express our thanks to Mr. Thaweedet Chainapong for collecting the specimens in 2002, and Mr. Taweesak Khuantrairong for his cooperation in field surveys in 2003-2004 and for preparing the illustrations.

REFERENCES

1. S. Kumano, "Freshwater Red Algae of the World", Biopress Ltd, Bristol, **2002.**
2. C. A. Agardh, "Systema algarum Lund XXXVIII", in "Freshwater Red Algae of the World" (Ed. S. Kumano), Biopress Ltd, Bristol, **2002.**
3. Y. Yamada, "On the species of *Thorea* from the far eastern Asia", *Jpn. J. Bot.*, **1949,** *24*, 155-159.
4. C. S. Lobban, F. K. Daily, W. A. Daily, R. W. Hoshaw and M. Schfter, "Periphyton, excluding diatoms and desmids, from Yap, Caroline Islands", *Micronesica*, **1990**, *23*, 27-40.
5. M. Rantasabapathy and R. Seto, "*Thorea prowsei* sp. nov. and *Thorea clavata* sp. nov. (Rhodophyta, Nemalinales) from West Malaysia", *Jpn. J. Phycol.*, **1981**, *29*, 243-250.
6. R. Seto, "Taxonomy and distribution of the taxa of Thoreaceae, freshwater Rhodophyta, in the western Pacific regions", Report of grant-in-aid for scientific researches from the Ministry of Education, Science and Culture, Japan, **1985** (in Japanese).
7. S. Kumano, T. Hanyuda, H. Sato, S. Traichaiyaporn, Y. Sugawa and S. Arai, "Morphology and phylogeny of the taxa of *Thorea* from the Asian Pacific Region", Proceedings of 4th Asian Pacific Phycological Forum, **2005**, Bangkok, Thailand, p.95.
8. S. Traichaiyaporn, T. Khuantrairong and S. Kumano, "*Thorea siamensis* sp. nov. (Thoreaceae: Rhodophyta) from Thailand", *Nat. Hist. J. Chulalongkorn Univ.*, **2008**, *8*, 27-33.

Maejo Int. J. Sci. Technol. **2012**, *6*(02), 224-237

Maejo International Journal of Science and Technology

ISSN 1905-7873
Available online at www.mijst.mju.ac.th

Full Paper

Fuzzy preference of multiple decision-makers in solving multi-objective optimisation problems using genetic algorithm

Surafel Luleseged Tilahun* and Hong Choon Ong

School of Mathematical Sciences, Universiti Sains Malaysia, 11800, Pulau Pinang, Malaysia

* Corresponding author: e-mail:surafelaau@yahoo.com

Received: 13 July 2011 / Accepted: 7 June 2012 / Published: 15 June 2012

Abstract: Most real-life optimisation problems involve multiple objective functions. Finding a solution that satisfies the decision-maker is very difficult owing to conflict between the objectives. Furthermore, the solution depends on the decision-maker's preference. Metaheuristic solution methods have become common tools to solve these problems. The task of obtaining solutions that take account of a decision-maker's preference is at the forefront of current research. It is also possible to have multiple decision-makers with different preferences and with different decision-making powers. It may not be easy to express a preference using crisp numbers. In this study, the preferences of multiple decision-makers were simulated and a solution based on a genetic algorithm was developed to solve multi-objective optimisation problems. The preferences were collected as fuzzy conditional trade-offs and they were updated while running the algorithm interactively with the decision-makers. The proposed method was tested using well-known benchmark problems. The solutions were found to converge around the Pareto front of the problems.

Keywords: multi-objective optimisation, multiple decision-makers, fuzzy preference, genetic algorithm

---

INTRODUCTION

A decision has to be made when there are a set of possible actions to choose from in order to optimise an objective or objectives. Decision-making is involved in almost all human experience. One can compare the outcome of a decision in order to choose from a set of possible actions. An action that yields the optimal outcome will be the best action to choose. An optimisation focuses on finding the best action for a given objective function or functions. If the number of objective functions is more than one, then the problem is called a multi-objective optimisation problem; otherwise, it is called a single-objective optimisation problem. Multi-objective optimisation problems are encountered in many real-life applications. They usually have conflicting objectives, such as maximising profit while improving the quality of the products. Hence, the solution to such problems involves a compromise between the objectives; this idea is known as the Pareto optimal solution after the Italian economist Vilfredo Pareto (1848–1923) [1]. A point in the feasible region for a multi-objective optimisation problem is said to be the Pareto optimal solution if it is not possible to find another feasible point which performs at least the same in all the objective functions and better in one or more of the objective functions [1]. There are usually many Pareto optimal solutions for a multi-objective optimisation problem. Choosing one from a set of Pareto solutions depends on the preference of the decision-maker. Ordering the objectives and trade-off are among the well-known and frequently used ways of expressing the preference of a decision-maker [2]. The trade-off can be expressed using fuzzy or crisp numbers. This preference helps to identify solutions according to the subjective judgment of the decision-maker [3].

Metaheuristic algorithms are useful in solving optimisation problems. Unlike classical approaches, these algorithms are not greatly affected by the behaviour of the problem. As a result, they have been widely used, especially in difficult optimisation problems. Although these algorithms do not guarantee optimality, they have been tested and proven to yield reasonable solutions [4]. Metaheuristic algorithms, especially genetic algorithms, have been used to solve many real problems formulated as multi-objective optimisation problems [5]. In a review paper, Coello [5] not only reviewed research trends in evolutionary algorithms for multi-objective optimisation problems, but also recommended further study on preference incorporation. Some studies of incorporating a decision-maker's preference involved the ranking of the objective functions [6, 7], which unfortunately lacks uniformity. Other studies involved the fuzzy trade-off of the decision-maker [8, 9] but did not consider situations involving multiple decision-makers or those where the trade-off varied from point to point in the feasible set. The impact of hedges, which uses a fuzzy trade-off as a preference, on fuzzy preferences has also not been dealt with in the previous papers. In many real-life problems, there may be multiple decision-makers with different decision-making powers [10].

In this paper, we introduce a genetic algorithm capable of embedding the fuzzy preferences of multiple decision-makers with different decision-making powers and solving multi-objective optimisation problems interactively. First, the fuzzy trade-off of each decision-maker is determined. From these trade-offs, a fuzzy weight for each decision-maker is constructed according to the randomly generated initial solutions. These solutions are incorporated in the fitness evaluation stage of the genetic algorithm by generating an appropriate probability density function. The probability density functions are constructed in such a way that they agree with the corresponding membership function of the fuzzy preferences and also with the hedges. After a specified number of iterations of the algorithm and depending on the solution at hand, a new preference is obtained, and the iteration is continued until a reasonable solution is achieved.

PRELIMINARIES

Multi-objective Optimisation

A multi-objective minimisation problem can be written as:

$$\min_{x \in S \subseteq \Re^n} F(x) = (f_1(x), f_2(x), \ldots, f_k(x)) \tag{1}$$

The problem is to find x^* in the feasible set S that yields a minimum value for the k objective functions f_i's, where $i = 1, 2, \ldots, k$. Usually, a conflict exists between the objective functions. Minimising one of the objective functions after some limits will lead to an increase in the value of the other objective functions.

Metaheuristic solution methods have become common in solving these type of problems. For instance, the evolutionary algorithm has been used in different studies [9, 11-13]. Although these solution methods do not guarantee optimality, they generate sound and acceptable solutions. These solution methods yield multiple solutions within a single run, in contrast to the classical solution methods, which rely on the conversion of multi-objective optimisation problems into single- objective optimisation problems, for example, the weighting method [14], Benson's method [14], the utility function method [3] and the lexicographic method [14, 15]. Choosing the best solution among a given set of solutions depends on the preference of the decision-maker. As stated by Coello [5], incorporating the decision-maker's preference is an important issue requiring further exploration.

Fuzzy Preference

A preference is a way to express the subjective judgment of the decision-maker. One commonly used way of determining preferences is using trade-offs. A conditional trade-off of objective j for a unit decrease of objective i at a given point, $(y_1, y_2, \ldots, y_i, \ldots, y_j, \ldots. y_k)$, is b, meaning that the decision-maker is indifferent between $(y_1, y_2, \ldots, y_i, \ldots, y_j, \ldots. y_k)$ and $(y_1, y_2, \ldots, y_i\text{-}1, \ldots, y_j+b, \ldots. y_k)$. This trade-off depends on the value $(y_1, y_2, \ldots, y_i, \ldots, y_j, \ldots. y_k)$. If we ask the decision-maker to make a trade-off from another point, it is possible to obtain another trade-off number different from b. Although it is common to use trade-off preference, it is not an easy task for the decision-maker. However, if the decision-maker is allowed the flexibility of assigning a trade-off fuzzily, it will make the task easier. This means that for a unit decrease of objective i, the decision-maker is willing to give around b units of objective j, yielding:

$$\left(y_1, y_2, \ldots, y_i, \ldots, y_j, \ldots. y_k\right) \sim \left(y_1, y_2, \ldots, y_i - 1, \ldots, y_j + t, \ldots. y_k\right), \text{ for } t \le b + d. \tag{2}$$

The symbol ~ in equation (2) indicates equivalency. As t becomes larger, the degree of acceptability or the equivalency continues to decrease and becomes unacceptable after exceeding some limit, e.g. after $t = b + d$. If d is defined as the width of the fuzzy interval and b as the average trade-off, it is possible to consider this acceptability of preference as a function of t as a membership function in the fuzzy set theory with t as a fuzzy number [16]. So, 'acceptable' has the membership value 1 and 'unacceptable' has the membership value 0. Generally, the values for the membership function, $z(t)$, which is between 0 and 1, should agree with the degree of acceptability, as shown in Figure 1.

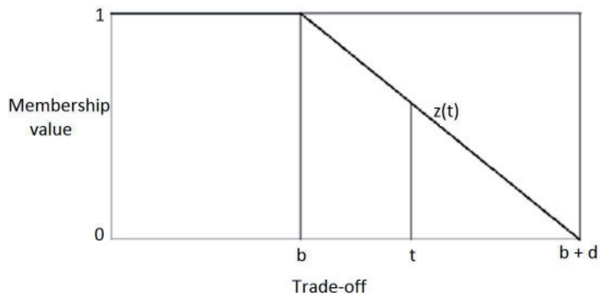


Figure 1. Fuzzy trade-off

Suppose that for the objective functions $f_i(x)$ and $f_j(x)$ with different i and j, the average trade-off given by the decision-maker is (b_{ij}) and the average width is (d_{ij}). Hence, we have two matrices: B = (b_{ij}) and D = (d_{ij}). It is meaningless to compute the trade-off of the i^{th} function for a unit decrease of the i^{th} function itself. Instead, we use b_{ii} = 0 and d_{ii} = 0. From B, it is possible to calculate the average weight, a weight with a high degree of acceptability, as follows:

$$\overline{b_p} = \frac{\sum_{\substack{i=1\\ i\neq p}}^{k} b_{pi}}{\sum_{j=1}^{k}\sum_{\substack{i=1\\ i\neq j}}^{k} b_{ij}} = \frac{\sum_{i=1}^{k} b_{pi}}{\sum_{j=1}^{k}\sum_{i=1}^{k} b_{ij}} \tag{3}$$

$$\overline{B} = \begin{pmatrix} \overline{b_1} \\ \overline{b_2} \\ . \\ . \\ . \\ \overline{b_k} \end{pmatrix} \tag{4}$$

where $\overline{B}$ is the average weight for the k objective functions.

It is also possible to compute the average normalised fuzzy width as follows:

$$\overline{d_p} = \frac{\left(\sum_{\substack{i=1\\ i\neq p}}^{k} d_{pi} \Big/ k-1 \right)}{\sum_{i=1}^{k}\left(\sum_{\substack{j=1\\ j\neq i}}^{k} d_{ij} \Big/ k-1 \right)} = \frac{\sum_{i=1}^{k} d_{pi}}{\sum_{i=1}^{k}\sum_{j=1}^{k} d_{ij}} \tag{5}$$

$$\overline{D} = \begin{pmatrix} \overline{d_1} \\ \overline{d_2} \\ . \\ . \\ . \\ \overline{d_k} \end{pmatrix} \tag{6}$$

where $\overline{D}$ is the normalised average fuzzy width.

For each function $f_i(x)$, the fuzzy weight, b_i, is around the corresponding average weight with some degree of acceptability, as shown Figure 2.

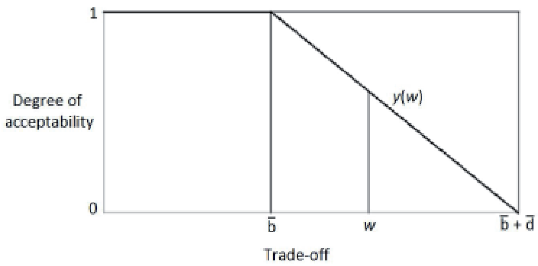
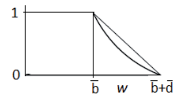
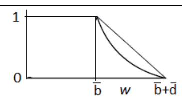
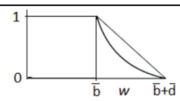
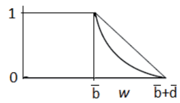
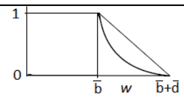
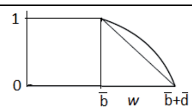
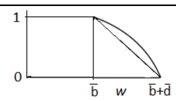
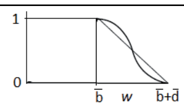


Figure 2. Fuzzy weight (where $\bar{b}$ stands for average weight of any of the objective functions, $\bar{b}_i$'s, and $\bar{d}$ is average width of any of the objective functions, $\bar{d}_i$'s)

Furthermore, the decision-maker may use hedges. Hedges are terms that modify the shape of the membership function of fuzzy sets. These include words such as very, somewhat, more or less, and slightly. The hedges, together with their graphical representations for our minimisation problem, are given in Table 1 and based on Negnevitsky [16].

Table 1. Graphical and functional representations of hedges for minimising the fuzzy trade-off

Hedges	The shape function (mathematical representation)	Graphical representation
A little	$h_{little}(w) = \left(y(w)\right)^{1.3}$	
Slightly	$h_{slightly}(w) = \left(y(w)\right)^{1.7}$	
Very	$h_{very}(w) = \left(y(w)\right)^{2}$	
Extremely	$h_{extremely}(w) = \left(y(w)\right)^{3}$	
Very very	$h_{very\_very}(w) = \left(y(w)\right)^{4}$	
More or less	$h_{more\_or\_less}(w) = \left(y(w)\right)^{1/2}$	
Somewhat	$h_{somewhat}(w) = \left(y(w)\right)^{1/2}$	
Indeed	$h_{indeed}(w) = \begin{cases} 2(y(w))^2, \text{if } 0 \le y(w) \le 0.5 \\ 1-2(1-y(w))^2, \text{if } 0.5 < y(w) \le 1 \end{cases}$	

Note: $y(w)$ is the straight line joining $(b,1)$ and $(b+d, 0)$

If $h_{ij}(w)$ is the mathematical representation of the hedge, when one collects the fuzzy conditional trade-off of f_j for a unit decrease of f_i, then $h_{ij}(w)$ can be any of the functional representations of the hedges in Table (1). If no hedges are used, then $h_{ij}(w)=y(w)$, which will be linear. Hence, there will be a matrix of functions, $H(w) = (h_{ij}(w))$:

$$H:\Re^k \rightarrow \Re^k \times \Re^k,\ H(w)=\begin{pmatrix} h_{11}(w) & h_{12}(w) & . & . & h_{1k}(w) \\ h_{21}(w) & . & . & . & h_{2k}(w) \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ h_{k1}(w) & . & . & . & h_{kk}(w) \end{pmatrix} \tag{7}$$

If $h_{ii}(w)$ is taken as the identity mapping for all i, it is possible to combine the given shape functions from the given hedges to obtain the average shape functions as follows:

$$\overline{h_p}(w)=\frac{\sum_{i=1}^{k} h_{pi}(w)}{\sum_{j=1}^{k}\sum_{i=1}^{k} h_{ij}(w)} \tag{8}$$

and where

$$\overline{h}(w)=\begin{pmatrix} \overline{h_1}(w) \\ \overline{h_2}(w) \\ . \\ . \\ . \\ \overline{h_k}(w) \end{pmatrix} \text{ is the average shape function.} \tag{9}$$

Hence, $\overline{h_p}(w)$ determines the shape of the fuzzy membership function for the p^{th} objective function with average weight $\overline{b_p}$ and average fuzzy width $\overline{d_p}$.

Genetic Algorithm

Heuristic algorithms have become useful for dealing with a wide range of problems [17-19]. Based on the evolutionary ideas of natural selection, genetic algorithms are a type of adaptive metaheuristic search algorithms used to find a solution to optimisation problems. The genetic algorithm used in this study is an evolutionary algorithm in which an initial set of solutions are generated. According to the fitness of the solutions, a selection is performed via crossover and mutation in order to construct a new population. The old population is updated by the fittest members, and the same process is continued until the termination criterion is fulfilled. The termination criterion can be the maximum number of generations (fixed number of iterations) or the stage when there is no more improvement in the fitness [20]. A flow chart for a genetic algorithm is given in Figure 3.

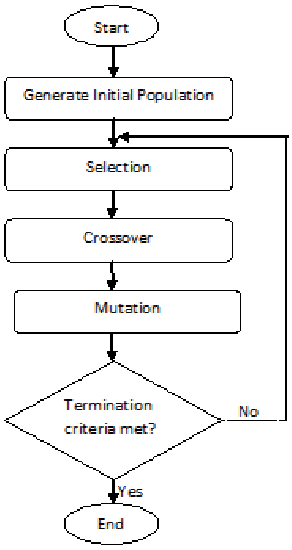


Figure 3. Flow chart of a genetic algorithm

MULTIPLE-PREFERENCE-INCORPORATED GENETIC ALGORITHM

Suppose there are m decision-makers, DM_1, DM_2, . . ., DM_m, with a specified authoritarian hierarchy dictating their preferences. DM_1 has the most authority and the level of authority decreases from DM_1 to DM_m. It is possible to assign a weight that describes the authority level of the decision-makers. One possible way to do this is to arrange the level of authority and assign a number which describes the level. We call it a vote. Let a vote be a numerical value which represents the degree of influence of the decision-maker in having his/her preference accepted. A decision-maker with a high level of authority will have a higher vote. When v_i represents the vote of DM_i, it is possible to construct a weight, c, for the decision-makers' preferences from the votes as follows:

$$c_i = \frac{v_i}{\sum_{j=1}^{m} v_j},\ i \in \{1,2,...,m\} \tag{10}$$

Suppose a decision-maker can give a cumulative fuzzy preference by looking at a set of feasible solutions. After using the fuzzy preference of each decision-maker to construct the average weight matrix, the average width matrix and the average fuzzy shape function for all decision-makers, it is possible to construct a total cumulative weight, W, a total cumulative width, R, and a cumulative shape function, $g(x)$, by combining $\overline{B}$, $\overline{D}$ and $\overline{h}(x)$ using the weights for the decision-makers.

For each decision-maker DM_i, we have:

$$\overline{B^i} = \begin{pmatrix} \overline{b^i_1} \\ \overline{b^i_2} \\ \cdot \\ \cdot \\ \cdot \\ \overline{b^i_k} \end{pmatrix},\quad \overline{D^i} = \begin{pmatrix} \overline{d^i_1} \\ \overline{d^i_2} \\ \cdot \\ \cdot \\ \cdot \\ \overline{d^i_k} \end{pmatrix} \quad \text{and} \quad \overline{h^i}(w) = \begin{pmatrix} \overline{h^i_1}(w) \\ \overline{h^i_2}(w) \\ \cdot \\ \cdot \\ \cdot \\ \overline{h^i_k}(w) \end{pmatrix}.$$

$$W = \frac{1}{m}\sum_{i=1}^{m} c_i \overline{B^i} \Rightarrow \begin{pmatrix} \overline{w}_1 \\ \overline{w}_2 \\ . \\ . \\ . \\ \overline{w}_k \end{pmatrix} = \begin{pmatrix} \frac{1}{m}\sum_{i=1}^{m} c_i \overline{b^i_1} \\ \frac{1}{m}\sum_{i=1}^{m} c_i \overline{b^i_2} \\ . \\ . \\ . \\ \frac{1}{m}\sum_{i=1}^{m} c_i \overline{b^i_k} \end{pmatrix} \tag{11}$$

$$R = \frac{1}{m}\sum_{i=1}^{m} c_i \overline{D^i} \Rightarrow \begin{pmatrix} r_1 \\ r_2 \\ . \\ . \\ . \\ r_k \end{pmatrix} = \begin{pmatrix} \frac{1}{m}\sum_{i=1}^{m} c_i \overline{d^i_1} \\ \frac{1}{m}\sum_{i=1}^{m} c_i \overline{d^i_2} \\ . \\ . \\ . \\ \frac{1}{m}\sum_{i=1}^{m} c_i \overline{d^i_k} \end{pmatrix} \tag{12}$$

$$g(w) = \frac{1}{m}\sum_{i=1}^{m} c_i \overline{h^i}(w) \Rightarrow \begin{pmatrix} g_1(w) \\ g_2(w) \\ . \\ . \\ . \\ g_k(w) \end{pmatrix} = \begin{pmatrix} \frac{1}{m}\sum_{i=1}^{m} c_i \overline{h^i_1}(w) \\ \frac{1}{m}\sum_{i=1}^{m} c_i \overline{h^i_2}(w) \\ . \\ . \\ . \\ \frac{1}{m}\sum_{i=1}^{m} c_i \overline{h^i_k}(w) \end{pmatrix} \tag{13}$$

Once we have the total cumulative weight, width and shape function, it is possible to incorporate these parameters in the fitness evaluation step of the genetic algorithm by generating a dynamic weight from the given preference and taking the weighted sum of the objective functions. To generate a random weight that satisfies the given cumulative preference, it is necessary to specify an appropriate probability density function for each dynamic weight, $g_i(w)$ for $i \in \{1, 2, . . ., k\}$, in such a way that a number with high acceptability needs to have a high probability, as shown in Figure 2. For instance, suppose no hedges are used, let the straight line joining $(\overline{w}_i, 1)$ and $(\overline{w}_i + r_i, 0)$ be $g_i(w) = p_i w + q_i$ for $i \in \{1,2, . . ., k\}$, for some p_i and q_i. It is necessary to generate a random weight under the shape function with a high probability near $\overline{w}_i$. For such purpose, it is possible to use a sampling method. To make the curve a probability density function, the area under the curve needs to be 1. To do so, it may be necessary to adjust the end points of the curve. Suppose it passes through $(\overline{w_i}, y)$ and $(\overline{w_i} + r_i, 0)$, where y is an arbitrary number that is determined by setting the area under the curve to 1:

$$\text{Area} = \int_{\overline{w_i}}^{\overline{w_i}+r_i} g_i(w)dw = 1 .$$

Furthermore, $g_i(\overline{w_i} + r_i) = 0$. Hence, $g_i(w)$ will be:

$$g_i(w) = \frac{-2w_i}{r_i^2} + \frac{2(\overline{w_i} + r_i)}{r_i^2}, \text{ for all } i. \tag{14}$$

In other words, w_i is a random variable with the probability density function $g_i(w)$. In a genetic algorithm, it is possible to incorporate this fuzzy preference and zoom to the area in the outcome space according to the cumulative fuzzy preference. To do that, we construct the fitness function in each iteration, j, by taking the weighted sum of the objective functions:

$$\text{Fitness function } = \sum_{i=1}^{k} w_i(j) f_i(x) \tag{15}$$

where $w_i(j)$ is the weight of f_i at iteration j.

The conditional fuzzy trade-off depends on the current value of the objective functions. Thus, it is necessary to return to the decision-makers after a number of iterations of the algorithm to determine whether a reasonable solution has been achieved. If the decision-makers are not satisfied with the solution, the preference will be updated by generating new preferences and rerunning the algorithm with new cumulative fuzzy preferences. Hence, the fuzzy weight may vary in the process until the decision-makers are satisfied with the solution or no further improvements can be made. This means that the algorithm will run interactively with the decision-makers until a termination criterion is fulfilled. The preference-incorporated interactive algorithm can be generalised as follows:

Initial step: set the parameter and the initial inputs as:

$f_j(x),\ x \in \Re^n,\ j \in \{1, 2, ..., k\}$ are the objective functions,

S is the feasibility condition,

P_r and P_m are the probability of crossover and mutation

and c_i is the vote of the decision-makers.

Main step:

1. $x_i, i \in \{1, 2, ..., m\}$, - generate the initial population and obtain a cumulative fuzzy preference from the DMs, depending on x_i's.
2. $g_i(w)$, - construct a probability density function for the weight of each objective function using c_i and the cumulative fuzzy preference.
3. $children = \varnothing$, - set the children set empty
   for *t=1* to the number of the generation
   $parent = \{x_i, i \in \{1, 2, ..., m\}\}$.
   (3.1) For the fitness function construction
   w_j $j \in \{1, 2, ..., k\}$, generate a weight for the objective functions using the probability density function $g_i(w)$.

$fun_i = \sum_{j=1}^{m} w_j f_j(x_i) \quad i \in \{1,2,...,m\}$, - dynamic fitness function

(3.2) Choose x_1' and x_2' for crossover and mutation with the probability related to the fitness.

(3.3) Apply crossover and mutation with the probability P_r and P_m, on x_1' and x_2' to obtain y_1' and y_2'.

$children = children \cup \{y_1', y_2'\}$, and update the set children.

If the number of elements in children < *m*, then return to step (3.2).

(3.4) Choose and set the fittest members as parents from $children \cup parents$

end.

4. If the decision-makers are satisfied with the solution, stop; otherwise, update the preference and go back to step (2).

EXPERIMENTAL RESULTS

Simulations using Matlab were carried out on five bi-objective optimisation problems, as in equation (16), with different Pareto fronts and different objective functions, f_i's. The preferences were changed twice. In all the tests, the probability of crossover and mutation were taken to be 0.9 and 0.2 respectively. Forty initial solutions were generated for all the simulations.

$$\min_{x \in S \subseteq \Re^2} F(x) = (f_1(x), f_2(x)) \tag{16}$$

The results of the simulations are presented below:

1. The first optimisation problem, given in equation (17), exhibited a convex Pareto front. The total average weight and width were changed once and the shape functions were linear. The total average weight and width were 0.6 and 0.15 respectively for the first objective function, and 0.5 and 0.2 respectively for the second. After 15 iterations, the parameters were changed to 0.3 and 0.15 for the first test function and 0.8 and 0.2 for the second respectively. The iteration number was taken to be 15 before and 15 after the change of preference. The results are shown in Figure 4.

$$f_1(x) = \frac{1}{2}\sum_{i=1}^{2} x_i^2, f_2(x) = \frac{1}{2}\sum_{i=1}^{2}(x_i - 2.0)^2 \text{ and } S = \{(x_1, x_2) \in \Re^2 \mid -1 \le x_1, x_2 \le 1\} \tag{17}$$

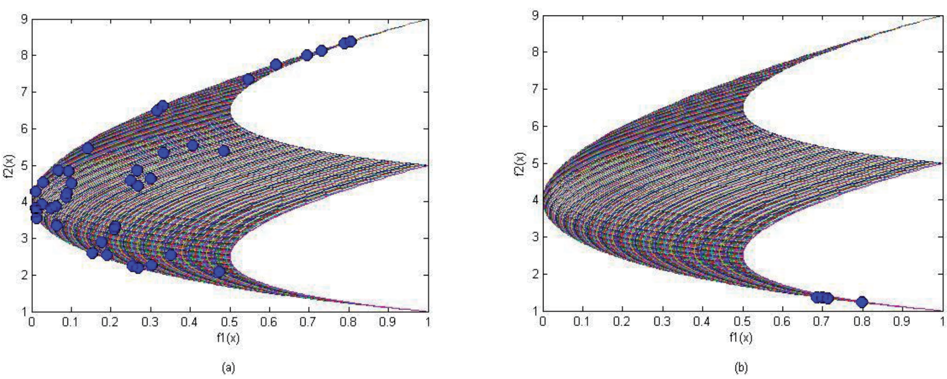


Figure 4. The outcome space for the first test problem: (a) with the population generated initially; (b) at the end of the algorithm

2. The second test problem exhibited a discontinuous Pareto front as shown in equation (18). The axis parallel to the hyper-ellipsoid function of the first objective function was negative, and the second objective function was a Rastrigin's function, with n = 2. The preferences were the same as in the first simulation, with the same number of iterations and shape functions. The number of iterations was 15 before and 15 after the change in the preferences. The results are shown in Figure 5.

$$\begin{aligned} &f_1(x) = -\left(2x_1^{\,2} + x_2^{\,2}\right), \\ &f_2(x) = 20 + x_1^{\,2} - 10\cos(2\pi x_1) + x_2^{\,2} - 10\cos(2\pi x_2) \\ &\text{and } -5 \le x_1, x_2 \le 5. \end{aligned} \tag{18}$$

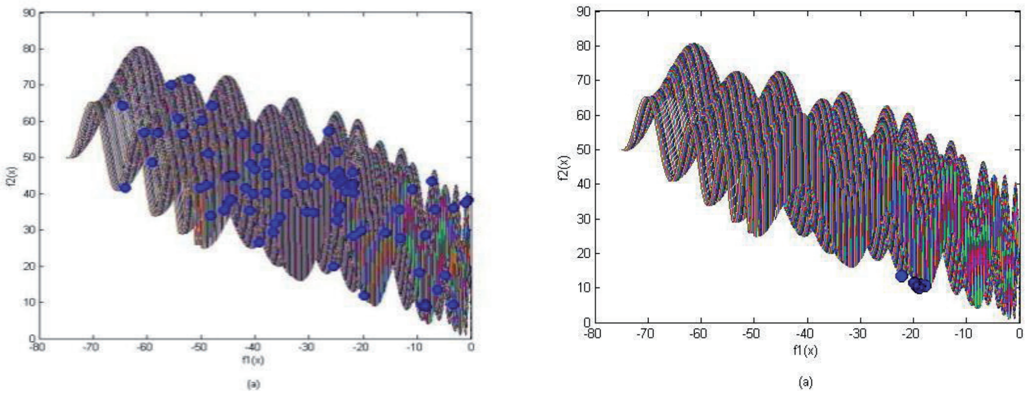


Figure 5. The outcome space for the second test problem: (a) with the population generated initially; (b) at the end of the algorithm

3. Here, the test problem had a point Pareto front. The first objective function was the first function of De Jong's, with n = 2, as shown in equation (19). The preferences were changed once with the same value used as in the previous cases, as well as the same shape functions and the same number of iterations. The results are shown in Figure 6.

$$f_1(x) = x_1^{\,2} + x_2^{\,2},\ f_2(x) = |\,x_1\,|^2 + |\,3x_2\,| \text{ and } 0 \le x_1, x_2 \le 1 \tag{19}$$

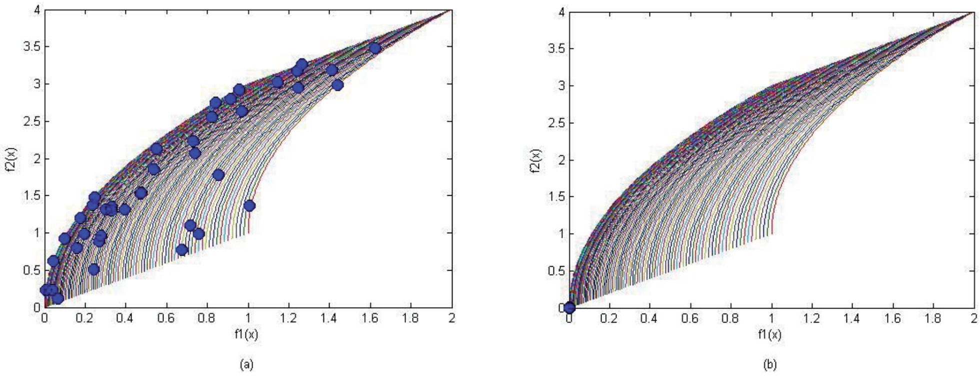


Figure 6. The outcome space for the third test problem: (a) with the population generated initially; (b) at the end of the algorithm

4. The fourth test problem exhibited a concave and convex Pareto front. The first objective function was a linear function (an identity function) as given in equation (20). The total average weight and width were 0.4 and 0.1 for the first objective functions, and 0.5 and 0.1 for the second respectively. These were changed after 15 iterations to 0.3 and 0.1 for the first objective functions, and 0.7 and 0.15 for the second respectively, with the shape functions left as a straight line. After changing the preferences, 15 iterations were performed. The results are shown in Figure 7.

$$f_1(x) = x_1,\ f_2(x) = (1+9x_2)\left(1 - (\frac{x_1}{1+9x_2})^{1/4} - (\frac{x_1}{1+9x_2})^4\right) \text{and } 0 \le x_1, x_2 \le 1. \tag{20}$$

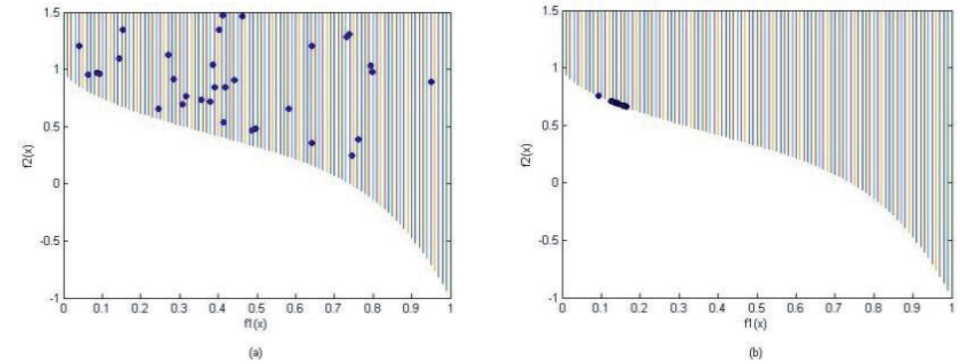


Figure 7. The outcome space for the fourth test problem: (a) with the population generated initially; (b) at the end of the algorithm

5. The last test problem showed a concave Pareto front as presented in equation (21). Fifteen iterations were performed before and after changing the preferences. The preferences and the shape functions were the same as in the previous case (test problem 4). The results are shown in Figure 8.

$$f_1(x) = x_1,\ f_2(x) = (1+9x_2)\left(1 - (\frac{x_1}{1+9x_2})^2\right) \text{and } 0 \le x_1, x_2 \le 1. \tag{21}$$

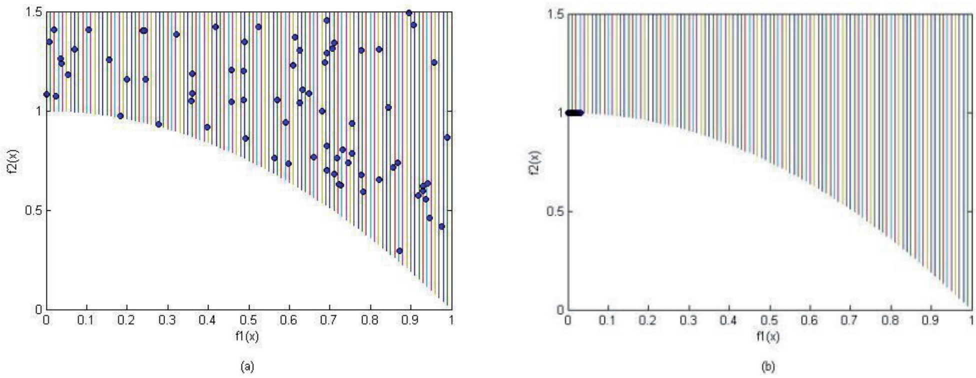


Figure 8. The outcome space for the fifth test problem: (a) with the population generated initially; (b) at the end of the algorithm

From the simulation results, one can see that the fuzzy preference incorporating an interactive genetic algorithm in this study is in agreement with the given preferences. When high preference is placed on the first objective function, f_1, the points move to the left; when preference for the objective function 2, f_2, is higher, the points move to the right.

CONCLUSIONS

Solving multi-objective optimisation problems using a genetic algorithm with varying fuzzy preferences of multiple decision-makers was discussed and demonstrated. The fuzzy trade-off preference of each decision-maker was determined from initially generated solutions. These preferences were combined so that fuzzy weights could be constructed according to the fuzzy trade-offs of the decision-makers. Using the vote, a numerical value, of each decision-maker in having his/her preference accepted, a cumulative fuzzy weight vector of dimension k was constructed. The cumulative fuzzy weight was then expressed using the probability density function, which was in agreement with the membership function and with the hedges given, if any. These cumulative weights were then incorporated with the genetic algorithm, specifically in the fitness evaluation stage. This interactive method was tested on five selected bi-objective optimisation problems with different Pareto fronts and properties. Based on the simulation, it is clear that the algorithm yields a solution along the Pareto front that is in agreement with the given preference.

ACKNOWLEDGEMENTS

This work was supported in part by Universiti Sains Malaysia short-term grant number 304/PMATHS/6311126. The first author acknowledges support from a TWAS-USM 2008 fellowship.

REFERENCES

1. M. Ehrgott and X. Gandibleux, "Multiple Criteria Optimisation: State of the Art Annotated Bibliographic Surveys", Kluwer Acadamic Publishers, Boston, **2002**.
2. L. Rachmawati and D. Srinivasan, "Preference incorporation in multi-objective evolutionary algorithms: A survey", Proceedings of IEEE Congress on Evolutionary Computation, **2006**, Vancouver, Canada, pp.962-968.
3. R. L. Keeney and H. Raiffa, "Decision with Multiple Objectives: Preferences and Value Tradeoffs", John Wiley and Sons, New York, **1976**.
4. X.-S. Yang, "Nature-Inspired Metaheuristic Algorithms", Luniver Press, Frome, **2010**.
5. C. A. Coello, "Evolutionary multi-objective optimization: Some current research trends and topics that remain to be explored", *Front. Comp. Sci. China*, **2009**, *3*, 18-30.
6. Y. Jin, M. Olhofer and B. Sendhoff, "Dynamic weighted aggregation for evolutionary multi-objective optimization: Why does it work and how?", Proceedings of Genetic and Evolutionary Computation, **2001**, San Francisco, USA, pp.1042-1049.
7. Y. Jin and B. Sendhoff, "Incorporating of fuzzy preferences into evolutionary multi-objective optimization", Proceedings of 4th Asia-Pacific Conference on Simulated Evolution and Learning, **2002**, Singapore, pp.26-30.

8. S. L. Tilahun and H. C. Ong, "Fuzzy preference incorporated evolutionary algorithm for multi-objective optimization", Proceedings of International Conference on Advanced Science, Engineering and Information Technology, **2011**, Bangi-Putrajaya, Malaysia, pp.26-30.
9. H. C. Ong and S. L. Tilahun, "Integrating fuzzy preference in genetic algorithm to solve multi-objective optimization problems" *Far East J. Math. Sci.*, **2011**, *55*, 165-179.
10. P. Jaramillo, R. A. Smith and J. Andreu, "Multi-decision-makers equalizer: A multi-objective decision support system for multiple decision-makers", *Ann. Operat. Res.*, **2005**, *138*, 97-111.
11. A. Lahsasna, R. N. Ainon and T. Y. Wah, "Enhancement of transparency and accuracy of credit scoring models through genetic fuzzy classifier", *Maejo Int. J. Sci. Technol.*, **2010**, *4*, 136-158.
12. C. L. Karr and L. M. Freeman, "Industrial Applications of Genetic Algorithms", CRC Press, Boca Raton, **1999**.
13. A. Osyczka and S. Krenich, "Evolutionary algorithms for multi-criteria optimization with selecting a representative subset of Pareto optimal solutions. Evolutionary multi-criterion optimisation", Proceedings of 1st International Conference on Evolutionary Multi-Criterion Optimisation, **2001**, Zurich, Switzerland, pp.141-153.
14. M. Ehrgott, "Multicriteria Optimization", Springer, Berlin, **2005**.
15. V. A. Emelichev, M. K. Kravtsov and O. A. Yanushkevich, "Lexicographic optima in the multicriteria discrete optimization problem", *Math. Notes*, **1995**, *58*, 928-932.
16. M. Negnevitsky, "Artificial Intelligence: A Guide to Intelligent System", Henry Ling, Harlow, **2005**.
17. M. Uysal, "Using heuristic search algorithms for predicting the effort of software projects", *Appl. Comput. Math.*, **2009**, *8*, 251-262.
18. K. Y. Lee and M. A. El-Sharkawi, "Modern Heuristic Optimization Techniques: Theory and Application to Power System", John Wiley and Sons, Hoboken, **2008**.
19. I. H. Osman and J. P. Kelly, "Meta-Heuristics: Theory and Applications", Kluwer Academic Publishers, Norwell, **1996**.
20. R. L. Haupt and S. E. Haupt, "Practical Genetic Algorithms", John Wiley and Sons, Hoboken, **2004**.

Maejo Int. J. Sci. Technol. **2012**, *6*(02), 238-248

Maejo International Journal of Science and Technology

ISSN 1905-7873
Available online at www.mijst.mju.ac.th

Full Paper

Simulation analysis of security performance of DPSK-OCDMA network via virtual user scheme

Vishav Jyoti * and Rajinder Singh Kaler

Department of Electronics and Communication Engineering, Thapar University, Patiala-147004, Punjab, India

* Corresponding author, e-mail: vishavjyoti@gmail.com

Received: 27 July 2011 / Accepted: 26 June 2012 / Published: 6 July 2012

Abstract: A novel technique to enhance the security of an optical code division multiple access (OCDMA) system against eavesdropping is proposed. It has been observed that when a single user is active in the network, an eavesdropper can easily sift the data being transmitted without decoding. To increase the security, a virtual user scheme is proposed and simulated on a differential phase shift keying (DPSK) OCDMA system. By using the virtual user scheme, the security of the DPSK-OCDMA system can be effectively improved and the multiple access interference, which is generally considered to be a limitation of the OCDMA system, is used to increase the confidentiality of the system.

Keywords: DPSK, eavesdropping, OCDMA

INTRODUCTION

The potential for enhanced data security is one of several advantages of the optical code division multiple access (OCDMA) [1]. If multiple codes operate simultaneously, it is almost impossible for an eavesdropper to get meaningful information because of multiple access interference (MAI) caused by all transmitting users. However, a single transmitting user is vulnerable to the eavesdropper's attack. At first, the notion that an OCDMA encoded signal is similar to a noise waveform makes it difficult for an eavesdropper to read the transmitting data. Nevertheless, the single-user on-off keying OCDMA (OOK-OCDMA) system can be easily attacked by a simple energy detector without any knowledge of the code [2, 3].

Therefore, the OCDMA system with a modulation format based on code shift keying (CSK) was introduced where data bits '0' and '1' were encoded by two different codes to improve its security against simple energy detection [2, 4]. However, the CSK-OCDMA system could be attacked by a differential phase-shift keying (DPSK) demodulator followed by a balanced photodetector [5]. To provide security against a standard power detector, a DPSK-OCDMA system was implemented [6]. However, eavesdroppers can still detect the data from the encoded signals by using a DPSK demodulator [5]. In this study, a DPSK-OCDMA network is considered to ensure data security against differential eavesdropping.

An unauthorised access is basically a threat to the network confidentiality when an individual user is isolated during transmission. An eavesdropper in an OCDMA network can isolate an individual user's signal from various locations within the network [7]. In a network where only a single user is transmitting, an eavesdropper can isolate the individual user's signal at a location within the network before multiplexing or coupling the multiple users' signals. An eavesdropper can also isolate an individual user's signal when a single user is active in the network while all other users are idle. In this paper, a novel technique is proposed and simulated to make the OCDMA system less vulnerable to eavesdropping by creating a virtual user environment in the DPSK-OCDMA network.

PROPOSED MODEL

In the OCDMA system, the data from all users is multiplexed before transmission onto an optical fibre. In the case when a single user is active in the network while all others are not transmitting, an eavesdropper can easily sift the data by tapping into an optical fibre. This problem can be solved by a virtual user scheme as shown in Figure 1. In this scheme, a virtual user is created that is always transmitting in parallel with the authorised users. Therefore, at the time when only a single user is transmitting and all other users are idle, the virtual user is the source of multiple access interference that makes the eavesdropper's task difficult. Pseudo random noise is given as the data input to the virtual user and the data are encoded using a unique optical code from the code set used. Further, this encoded stream is multiplexed with the other user's data stream before transmitting the signal onto the optical fibre. This virtual user will serve as an interferer because it appears as an authorised user to an eavesdropper, thus hindering eavesdropping. Therefore, an eavesdropper can never isolate an individual user's signal because there is always an active virtual user in the network that serves as interferer, and deciphering the data bits when multiple users are present is more difficult [8]. In an OCDMA network, MAI is generally considered as a limitation but in a virtual user OCDMA system, it can be advantageously used to increase the confidentiality of the system.

This scheme was already analysed with OOK-OCDMA [9]. It was seen that irrespective of the encoding scheme, OCDMA system with OOK modulation format is not secure. A prototype virtual user scheme was proposed to enhance the security of OOK-OCDMA against simple energy detector when a single user is active in the network [9]. This proposed virtual user scheme increased security but imposed a bandwidth constraint on the system because a virtual user was incorporated with each user.

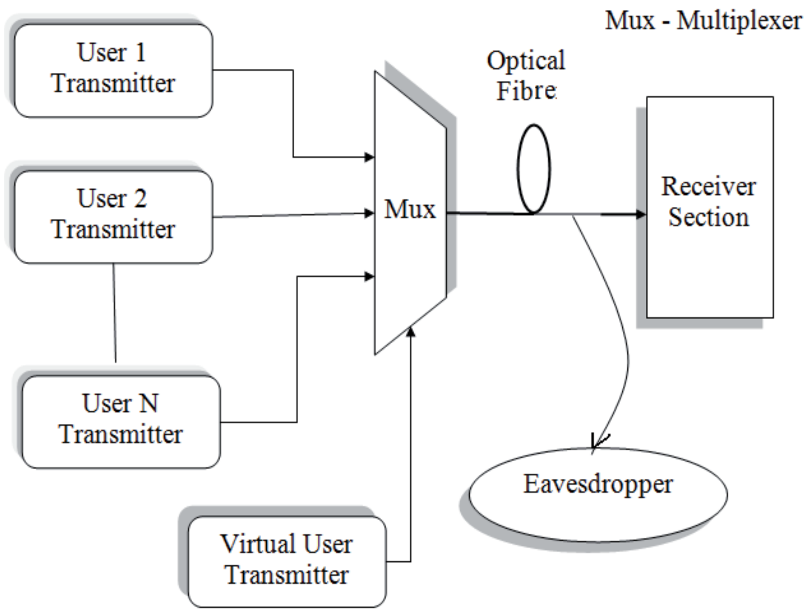


Figure 1. Proposed virtual user environment model for OCDMA network

In this article, the virtual user scheme is analysed with DPSK-OCDMA, which is vulnerable to a differential eavesdropper. Here, we propose a bandwidth-efficient virtual user scheme where all the authorised users have a common virtual user. The proposed scheme helps to make the system inherently secure against eavesdropping without affecting the system performance or imposing any additional bandwidth penalty.

The bit error rate (BER) for DPSK-OCDMA can be evaluated for both the single user and virtual user cases as given below [10]. For single-user OCDMA network, BER becomes:

$$BER_{single} = \frac{1}{4}\left(1 + erf\left(\frac{-A_g w}{\sqrt{2(\sigma_{th}^2)}}\right) + erfc\left(\frac{A_g w}{\sqrt{2(\sigma_{th}^2)}}\right)\right)$$

where w is the code weight, A_g is the peak power of a single chip and σ_{th}^2 is the thermal noise. The erf(x) is the error function defined as:

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt \qquad -\infty < x < \infty$$

and erfc(x) is the complementary error function:

$$\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$$

For virtual user with single user, BER becomes:

$$BER_{virtual} = \frac{1}{4}\left(1 + erf\left(\frac{-A_g w - A_g \frac{w^2}{2L}}{\sqrt{2\sigma_1^2}}\right) + erfc\left(\frac{A_g w + A_g \frac{w^2}{2L}}{\sqrt{2\sigma_1^2}}\right)\right)$$

where

$$\sigma_1^2 = A_g^2 \frac{w^2}{2L}\left(1 - \frac{w^2}{2L}\right) + \sigma_{th}^2$$

and L is the code length. Thus,

$$BER_{virtual} = \frac{1}{4}\left(1 + erf\left(\frac{-A_g w - A_g \frac{w^2}{2L}}{\sqrt{2\left(A_g^2 \frac{w^2}{2L}\left(1 - \frac{w^2}{2L}\right) + \sigma_{th}^2\right)}}\right) + erfc\left(\frac{A_g w + A_g \frac{w^2}{2L}}{\sqrt{2\left(A_g^2 \frac{w^2}{2L}\left(1 - \frac{w^2}{2L}\right) + \sigma_{th}^2\right)}}\right)\right)$$

This scheme has certain advantages over the code switching scheme since, for the same number of users, twice as many codes are needed in the code switching scheme and the increase in the number of codes adds complexity to the network, which may also increase the cost of network management [7]. On the other hand, unlike the case for the code switching scheme, the introduction of one virtual user would not affect the system performance because the number of authorised users is not halved with respect to a standard OCDMA transmission.

To illustrate this, the information capacity (C) of an OCDMA network may be given as [11]: $C = K. [\ 1 - \log_2 (1+e^{-SNR})\ ]$, where K = number of users and SNR is the signal to noise ratio which depends on MAI noise. For the code switching scheme, the number of users is halved because two codes are used in encoding for each user. Therefore, the system capacity is halved. For the proposed scheme, however, only one user is added to form a virtual user environment and this does not affect the system performance.

SIMULATION SET-UP

The security enhanced DPSK-OCDMA system implementing a virtual user environment is shown in Figure 2. An OCDMA system based on spectral coding with a DPSK modulation format for a single transmitting user was simulated using the OptSim software. The simulation parameters considered for DPSK-OCDMA are given in Table 1. Here, the case when a single user is transmitting while all other are idle and the virtual user is active in the network was considered.

First, an OCDMA system with DPSK signalling and balanced detection was simulated. The DPSK system with balanced detection is an attractive modulation format for long-haul transmission as compared to the OOK modulation [12,13]. The DPSK format is a subset of the phase-shift keying (PSK) format in which the information carrying part for the DPSK encoded data is the phase difference applied to the carrier corresponding to two consecutive data bits. If the previous bit is 0, no phase shift is applied for encoding the current bit. If the previous bit is 1, the phase of the carrier for the current bit is applied with a phase shift of 180 degrees [14].

Next, a virtual user was created which was always transmitting in parallel to the authorised user. The outputs of the authorised and virtual users were combined by the optical multiplexer before transmission. A conventional single mode optical fibre having a length of 25 km was used for transmission, after which the DPSK modulated signal entered the receiver section of the topology.

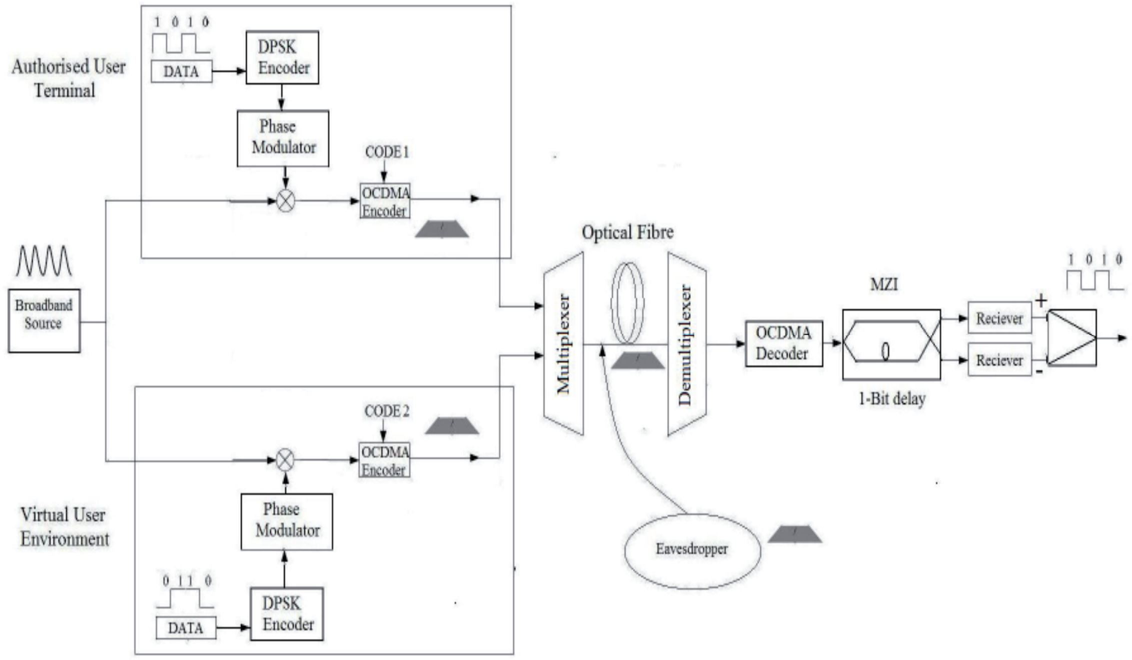


Figure 2. Security enhanced DPSK-OCDMA system with virtual user environment for single transmitting user

Table 1. Simulation parameters

Parameter	Value
Bit rate (Data rate)	2.5 Gbps
Number of optical sources	12 mode-locked lasers
Wavelength range	1550–1552.2 nm
Wavelength spacing	0.2 nm
Repetition rate of laser pulses	2.5 Gbps
Input power	1 –5 mW
Drive type of electric generator	On-off
Signal type of electric generator	Voltage
Modulation type	Phase modulation (with a phase shift of π)
Codes used	Zero cross correlation codes (ZCC)
Code weight	3
Code length	12
Fibre length	25 km
Attenuation	0.25 dB/km
Amplifier gain	30 dB
Delay in MZI	0.4 ns (1-bit delay)

At the receiver, a spectrally encoded signal was decoded by an authorised user sharing the same zero-cross correlation (ZCC) code [15,16] with the transmitter. One of the consequences of the DPSK format was that optical intensity remained constant during all bits, and thus the direct-detection receivers could not be used to detect the PSK signals. The demodulator consisted of a delay line interferometer that had a 1-bit delay; therefore, 2 bits could be compared at one time [17]. Differential time delay was set to the bit duration. Two outputs of the interferometer corresponded to 'constructive port' and 'destructive port' where maximum power appeared at the former when there was no phase change between adjacent bits, and at the latter when the phase in adjacent bits differed by π. Then two outputs of the demodulator were input to a balanced receiver that transformed the optical field into an electric current. Next, the output electrical signal from one of the receivers was inverted and both electrical signals were combined by an electrical summer.

After transmission, an eavesdropper employing a simple energy detector and a differential detector was placed. At the differential detector, the incoming signal was split into two paths and combined again with 1-bit difference between the two paths followed by a balanced photo detector [17].

RESULTS AND DISCUSSION

For both the DPSK-OCDMA and virtual user DPSK-OCDMA schemes, eye diagrams, BER and signals at different locations in the network were measured. The input signal of an authorised user is shown in Figure 3.

The input spectrum consisted of 12 wavelengths. After applying the ZCC code, only three wavelengths passed through the encoder as shown in Figure 4. The code weight was 3 as justified by the encoded spectrum and the code length was 12 as justified by the input spectrum.

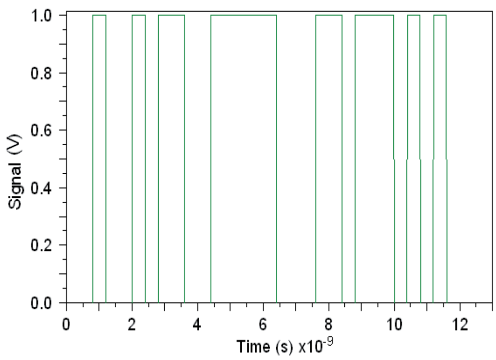
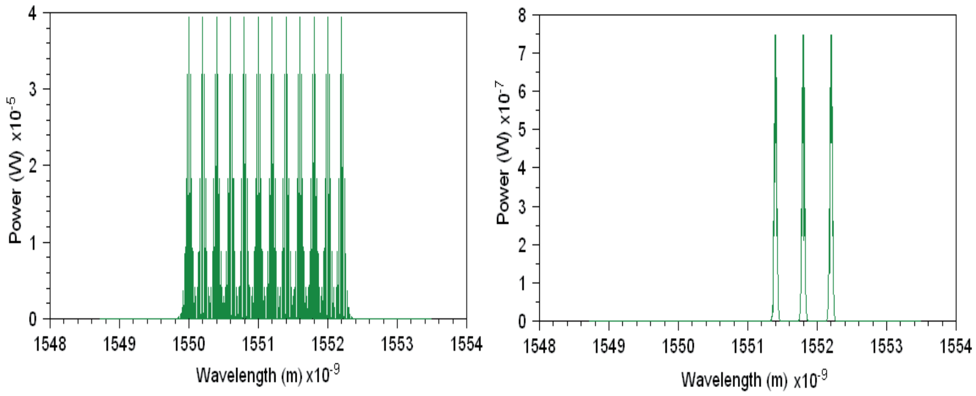


Figure 3. Input signal of authorised user

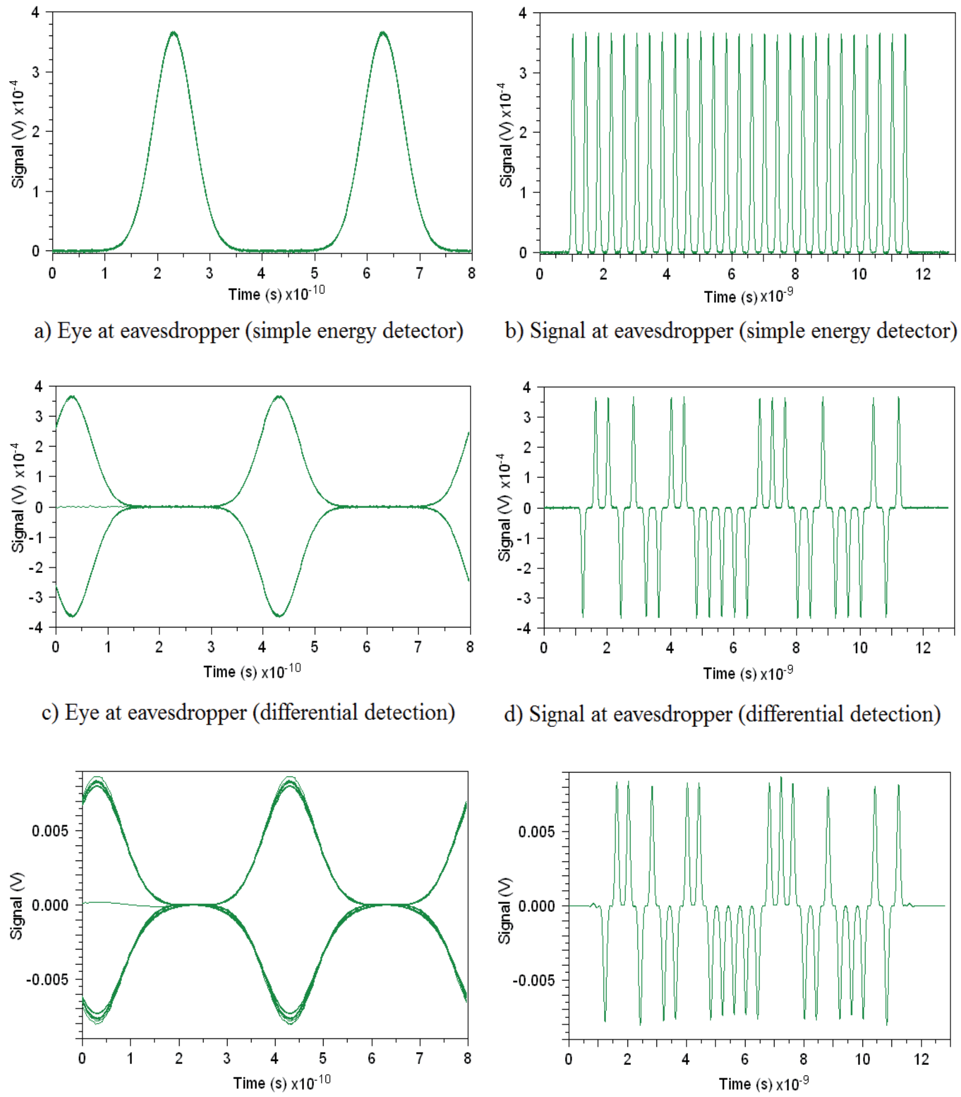
Figure 5(a) shows an eye diagram at the eavesdropper using a simple energy detector when only one user was transmitting. There was no eye opening (which is required) and no intelligible signal was present at the eavesdropper, as shown in Figure 5(b). There was no eye at all because the optical intensity remained constant during all bits, which demonstrates the ability to enhance security by using the DPSK-OCDMA system [4]. At the differential eavesdropper, differential detection was simulated directly without decoding. In this case, the differential detector clearly decoded the data signal without even knowing the code. A clear eye diagram was observed for the DPSK



a) Input Spectrum

b) Encoded Spectrum

Figure 4. Wavelength spectrum before and after encoding



a) Eye at eavesdropper (simple energy detector)

b) Signal at eavesdropper (simple energy detector)

c) Eye at eavesdropper (differential detection)

d) Signal at eavesdropper (differential detection)

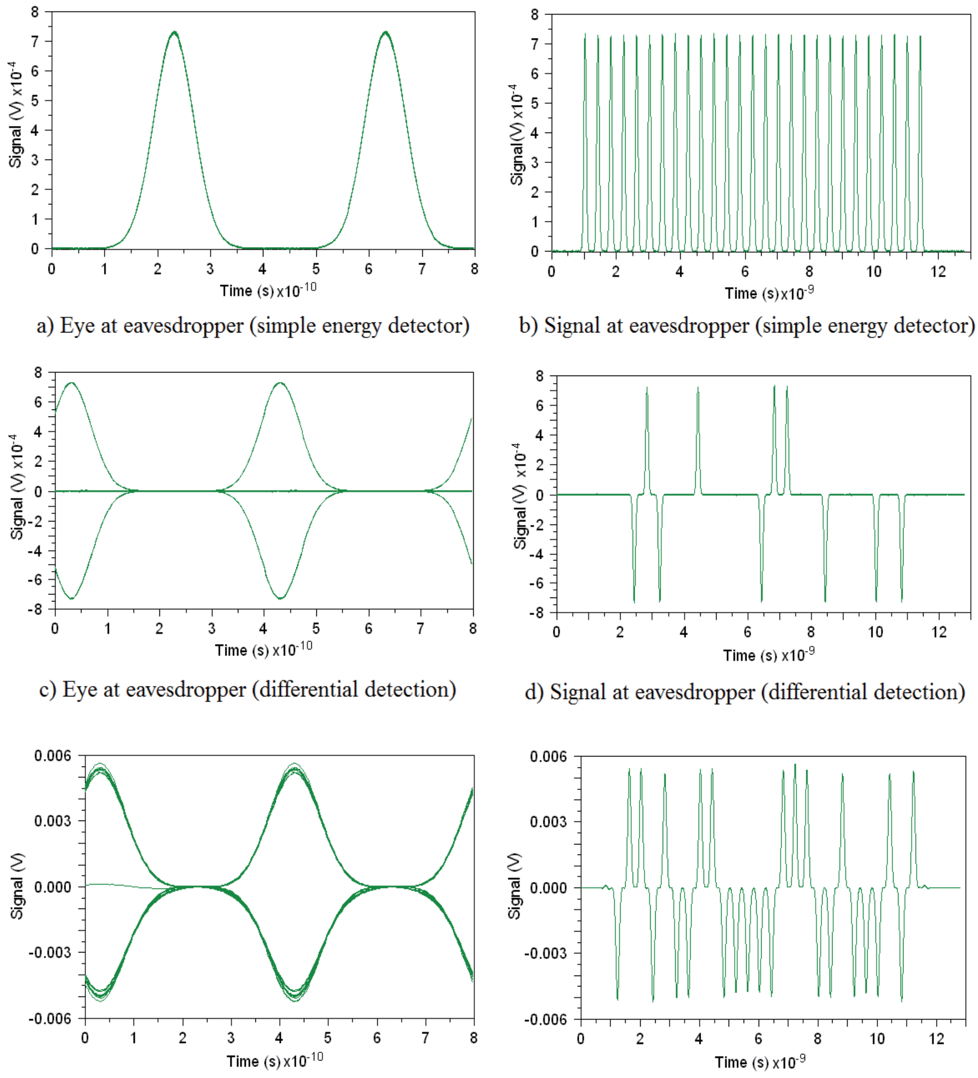
e) Eye diagram at receiver

f) Received signal

Figure 5. Eye diagrams and signals at various points when a single user is transmitting in DPSK-OCDMA network

eavesdropper as shown in Figure 5(c), and the detected signal waveform is shown in Figure 5(d). At the authorised receiver, differential detection was simulated after the OCDMA decoder. In Figure 5(e) also, a clear eye diagram is observed. The received signal is shown in Figure 5(f). It can be seen that the signals at the DPSK eavesdropper and the receiver completely overlap each other. This indicates that an eavesdropper can easily intercept the transmitted information by using differential detection. Therefore, the DPSK-OCDMA system is not secure for a single transmitting user in the presence of differential eavesdropping.

For the virtual user scheme where one user was always transmitting in parallel to the authorised user asynchronously, no eye was observed at the simple power detector, as shown in Figure 6(a), and the optical signal had only '1' bit as shown in Figure 6(b). Therefore, the data could



a) Eye at eavesdropper (simple energy detector)
b) Signal at eavesdropper (simple energy detector)
c) Eye at eavesdropper (differential detection)
d) Signal at eavesdropper (differential detection)
e) Eye diagram at receiver
f) Received signal

Figure 6. Eye diagrams and signals at various locations when single user is transmitting in virtual user DPSK-OCDMA network

not be detected by the direct detection receiver. Figures 6(c) and 6(d) show the eye diagram and detected signal respectively at the differential eavesdropper. The signal present at the DPSK eavesdropper did not overlap the received signal (Figure 6(f)), which indicates that the differential eavesdropper was getting a false data sequence instead of the original one. Figure 6(e) shows that a clear eye diagram was observed at the authorised receiver. The virtual user here served as an interferer making it difficult for an eavesdropper to properly decode the signal even when a single user was active in the network.

Further, Figure 7 shows the variation of BER as a function of input power at 2.5 Gbps for the DPSK-OCDMA network. It can be seen that eavesdropping with the differential detector can be harmful for this scheme as the BER varies from 10^{-9} to 10^{-11} with an increase in input power. This level of BER at the eavesdropper is sufficient to detect the transmitted signal correctly and the user security is compromised. Hence the above scheme is susceptible to differential eavesdropping. On the other hand, eavesdropping in the presence of the virtual user scheme gives a high value of BER regardless of the input power. Although the virtual user scheme gives a high BER at the eavesdropper, the BER obtained at the authorised receiver is below the threshold level (dotted red line) at all time and thus the information is successfully conveyed without compromising the security. So it is deduced that implementing the OCDMA system with a virtual user scheme ensures high security against eavesdropping with differential detection without affecting the received signal.

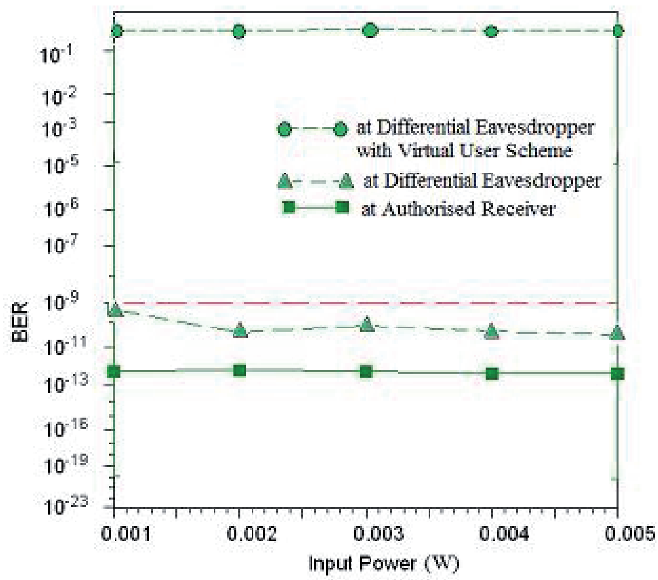


Figure 7. BER vs. input power in DPSK-OCDMA network with virtual user scheme

CONCLUSIONS

The virtual user scheme developed in this work increases the confidentiality of the DPSK-OCDMA system, the system in which a differential eavesdropper can easily read the information transmitted by a single user. The proposed scheme never allows an eavesdropper to isolate a single user's signal while all the other users are dormant or idle. A virtual user alongside an actual user in an OCDMA system mimics multiple access interference to make eavesdropping difficult. It is clearly seen from the results that the proposed virtual user scheme outperforms the conventional DPSK-OCDMA system in terms of enhanced security against differential eavesdropping. The scheme clearly increases the confidentiality of an OCDMA network without affecting the system performance or incurring any additional bandwidth penalty as compared to DPSK-OCDMA system. Moreover, as compared to the previous virtual user scheme, it is more bandwidth efficient since a common virtual user is created for all the authorised users.

REFERENCES

1. T. H. Shake, "Security performance of optical CDMA against eavesdropping", *IEEE J. Lightwave Technol.*, **2005**, *23*, 655-670.
2. D. E. Leaird, Z. Jiang and A. M. Weiner, "Experimental investigation of security issues in OCDMA: A code-switching scheme", *Electron. Lett.*, **2005**, *41*, 817-819.
3. Z. Jiang, D. S. Seo, S. D. Yang, D. E. Leaird, R. V. Roussev, C. Langrock, M. M. Fejer and A. M. Weiner, "Four-user, 2.5-Gb/s, spectrally coded O-CDMA system demonstration using low-power nonlinear processing", *IEEE J. Lightwave Technol.*, **2005**, *23*, 143-158.
4. X. Wang, N. Wada, T. Miyazaki, G. Cincotti and K. Kitayama, "Asynchronous multiuser coherent OCDMA system with code-shift-keying and balanced detection", *IEEE J. Select. Topics Quantum Electron.,* **2007**, *13*, 1463-1470.
5. B. Dai, Z. Gao, X. Wang, N. Kataoka and N. Wada, "Experimental investigation on security of temporal phase coding OCDMA system with code-shift keying and differential phase-shift keying", Proceedings of Asia Communications and Photonics Conference and Exhibition (ACP), **2010**, Shanghai, China, pp.427-428.
6. X. Wang, N. Wada, T. Miyazaki and K. Kitayama, "Coherent OCDMA system using DPSK data format with balanced detection", *IEEE Photonics Technol. Lett.*, **2006**, *18*, 826-828.
7. P. R. Prucnal, "Optical Code Division Multiple Access: Fundamentals and Applications", CRC Press, Boca Raton, **2006**.
8. Z. Jiang, D. E. Leaird and A. M. Weiner, "Security issues in OCDMA with multiple-user aggregation", Proceedings of Conference on Lasers and Electro-optics (CLEO), **2007**, Baltimore, USA, pp.1-2.
9. V. Jyoti and R. S. Kaler, "A novel virtual user scheme to increase data confidentiality against eavesdropping in OCDMA network", *Chin. Opt. Lett.*, **2011**, *9*, 120602.
10. G. Manzacca, F. Benedetto, V. Sacchieri, G. Giunta and G. Cincotti, "Advanced modulation formats in optical code division multiple access networks", Proceedings of International Conference on Transparent Optical Networks (ICTON), **2007**, Rome, Italy, pp.91-94.
11. G. Cincotti, N. Kataoka, N. Wada and K. Kitayama, "Perspectives of optical coding/decoding techniques in OCDMA networks", Proceedings of Asia Communications and Photonics Conference and Exhibition (ACP), **2009** , Shanghai, China, pp.1-2.
12. X. Wang, N. Wadat, T. Miyazaki and K. Kitayama, "Demonstration of DPSK-OCDMA with balanced detection to improve MAI and beat noise tolerance in OCDMA system", Proceedings

of Optical Fiber Communication Conference and National Fiber Optic Engineers Conference (OFC/NFOEC), **2006**, Anaheim, USA.

13. X. Wang, N. Wada, T. Miyazaki, G. Cincotti and K. Kitayama, "Advanced modulation techniques in OCDMA system", Proceedings of Optical Fiber Communication and Optoelectronics Conference, **2007**, Shanghai, China, pp.100-102.
14. J. S. Chitode, "Principles of Communication", Technical Publications, Pune, **2009**, pp.12-18.
15. M. S. Anuar, S. A. Aljunid, R. Badlishah, N. M. Saad and I. Andonomic, "Performance analysis of optical zero cross correlation in OCDMA system", *J. Appl. Sci.*, **2007**, *7*, 3819-3822.
16. V. Jyoti and R. S. Kaler, "Design and performance analysis of various one dimensional codes using different data formats for OCDMA system", *Optik-Int. J. Light Electron Opt.*, **2011**, *122*, 843-850.
17. B. Dai, Z. Gao, X. Wang, N. Kataoka and N. Wada, "Demonstration of differential detection on attacking code-shift-keying OCDMA system", *Electron. Lett.*, **2010**, *46*, 1680-1682.

Maejo Int. J. Sci. Technol. **2012**, *6*(02), 249-258

Maejo International Journal of Science and Technology

ISSN 1905-7873

Available online at www.mijst.mju.ac.th

Full Paper

Comparative properties of natural rubber vulcanisates filled with defatted rice bran, clay and calcium carbonate

Darinya Moonchai*, Natthapong Moryadee and Nipon Poosodsang

Rubber and Polymer Technology Programme, Faculty of Engineering and Agricultural Industry, Maejo University, Chiangmai 50290, Thailand

*Corresponding author, e-mail: darinya@mju.ac.th

Received: 8 April 2011 / Accepted: 12 July 2012 / Published: 13 July 2012

Abstract: The effects of defatted rice bran (DRB) as a filler for natural rubber vulcanisate on its cure characteristics, mechanical properties and morphology were investigated. The properties of the DRB-filled vulcanisate were also compared with clay-filled and $CaCO_3$-filled vulcanisates. At similar loading level (50 parts per hundred of rubber), DRB-filled vulcanisate gave the shortest cure time. Clay-filled vulcanisate showed highest tensile and tear strength followed by DRB-filled vulcanisate. However, $CaCO_3$-filled vulcanisate gave highest rebound resilience while DRB-filled vulcanisate exhibited highest modulus, hardness and abrasion resistance. Scanning electron micrographs revealed that the morphology of clay-filled vulcanisate was more homogenous than that of DRB-filled and $CaCO_3$-filled vulcanisates. According to these observations, DRB can potentially be used as a cheap and more environment-friendly natural filler when an improvement in mechanical properties was not so critical.

Keywords: natural rubber, vulcanisate, rubber filler, defatted rice bran

INTRODUCTION

Fillers are compounding ingredients added to rubber compounds for the purpose of reinforcing them and/or cheapening their cost. Traditional fillers include carbon black, silica, calcium carbonate, calcium silicate and clay. Carbon black is the most popular filler added to the rubber compounds due to its ability to enhance certain properties, especially mechanical properties [1-3]. Clay and calcium carbonate are considered as useful fillers in rubber compounds because of their low cost. Much work [4-7] has been done to study the use of clay and calcium carbonate as fillers for rubber compounds. Apart from the traditional fillers, the use of renewable materials such as starch, rice husk, rice husk ash, wood sawdust, rubber wood and soy meal have also appeared in the literature [8-18]. Fillers derived from renewable materials have attracted interest because of their low cost, renewability and environment-friendly nature.

Rice bran is a by-product of the rice-milling process. It is also the source of high-quality edible oil (rice bran oil). Rice bran oil is extracted from rice bran, leaving defatted rice bran (DRB) as by-product. DRB is used to reduce the final cost of animal feed or is discarded as agricultural waste. However, it still contains significant amounts of protein, carbohydrate, dietary fibre and phenolic substances [19]. To upgrade the value of DRB, we explore its possible application as a filler in rubber compounds. The effects of DRB on the morphology, cure characteristics and mechanical properties of vulcanised natural rubber (NR) are investigated in this paper. The properties of the DRB-filled vulcanisate are also compared with clay-filled and $CaCO_3$-filled vulcanisates.

MATERIALS AND METHODS

Materials

All materials were used as received. NR (STR 5L), elemental sulphur (S_8), stearic acid, zinc oxide, accelerators and antioxidant (polymer of p-cresol; Lowinox® CPL) were purchased from Lucky Four Co. Ltd. (Thailand). Two types of accelerators used were dibenzothiozyl disulphide (Vulkacit® MBTS) and tetramethylthiuram disulphide (Vulkacit® TMTD) (97% purity). Stearic acid and zinc oxide were of rubber-grade. Three fillers used were DRB, china clay (hydrated aluminium silicate) and ground calcium carbonate ($CaCO_3$). Clay and calcium carbonate were purchased from Lucky Four Co. Ltd. and DRB (composition shown in Table 1) was purchased from Thai Edible Oil Co. Ltd.

Table 1. Composition of DRB

Component	%
Crude protein[1]	17.57
Crude fibre[1]	9.19
Moisture content[1]	11.10
Crude fat and oil[1]	1.40
Carbohydrate[1]	49.74
Ash[2]	11.00

[1] Laboratory of Animal Nutrition Research and Development Centre, Kasetsart University
[2] Laboratory of Dairy Research and Development Centre, Kasetsart University

DRB was passed through a 150-mesh screen and dried in a circulating air oven at 70°C for 17 hr before mixing. The Brunauer-Emmett-Teller (BET) surface area and porosity of the fillers were determined with a nitrogen adsorption instrument (Quantachrome Autosorb-1, Quantachrome Corp., USA) according to ISO 9277 [20]. The samples were degassed at 100°C for 24 hr in the degas pot of the adsorption instrument in order to remove moisture and other contaminants before measurements, which were done at the boiling point of nitrogen (-196°C). The amount of adsorbed gas was measured to determine the surface area and porosity of the surface. The calculation was based on the BET theory [20]. The results are given in Table 2.

Table 2. Physical properties of different types of fillers

Filler	Specific surface area (m^2/g)	Pore volume (cc/g)	Pore diameter (Å)
DRB	4.63	0.0074	64.09
Clay	5.83	0.0182	124.80
$CaCO_3$	4.46	0.0136	122.00

Preparation of Rubber Compounds and Vulcanisates

All rubber compounds contained the same chemical composition except for the filler type. For each rubber compound, 50 parts per hundred of rubber (phr) of the filler content was used. The ingredients used in each compound are listed in Table 3. The four compound formulations are designated as control (no filler), DRB, clay and $CaCO_3$. The mixing was carried out both in an internal mixer (model YFD-3L, Yong Fong Machinery Co. Ltd., Thailand) and a two-roll mill (model YFTR-8, Yong Fong Machinery Co. Ltd., Thailand). All ingredients except sulphur were mixed with the rubber in the internal mixer with a fill factor of 0.7 at 80°C and a rotor speed of 50 rpm. The mixing sequence is shown in Table 4. After discharging, the compounds were further masticated in the two-roll mill for 2 min. Then sulphur was added and mixed with the rubber compounds for 3 min. Finally, the rubber compounds were taken out and sheeted through a two-roll mill. The rubber compounds were compression-moulded at 150°C using a hydraulic hot press (OOMN semi-automatic moulding press model HPC-100(D), Shanghai Zimmerli Weili Rubber and Plastic Machinery Co. Ltd., China) according to their respective cure time (t_{90}) from the cure curves.

Table 3. Formulations of rubber compounds

Ingredient	Amount (phr)			
	Control	DRB	Clay	$CaCO_3$
NR (STR 5L)	100	100	100	100
Sulphur	2.5	2.5	2.5	2.5
Stearic acid	2	2	2	2
Zinc oxide	4	4	4	4
MBTS	1	1	1	1
TMTD	0.5	0.5	0.5	0.5
CPL	1	1	1	1
DRB	-	50	-	-
Clay	-	-	50	-
$CaCO_3$	-	-	-	50

Note: phr = parts per hundred of rubber

Table 4. Mixing sequence of ingredients in the internal mixer

Minute of addition	Operation
0	Loading of NR
1	Mastication of NR
5	Addition of stearic acid
6	Addition of zinc oxide
7	Addition of half of filler
9	Addition of MBTS, TMTD and CPL
11	Addition of rest of filler
13	Discharging

Note: For control blend, total mixing time was 9 min. (no filler addition)

Cure Characteristics

The cure characteristics of the different rubber compounds were evaluated using a moving die rheometer (model UR-2010, U-CAN Dynatex Inc., Taiwan) which was operated at 150°C with 3° arc for 60 min., following ISO 6502 [21]. Minimum torque (ML), maximum torque (MH), scorch time (t_s) and cure time (t_{90}) were determined. The cure time, the time at which the rheometer torque increases to 90% of the total torque change on the cure curve, was obtained from the moving-die rheometer. The cure characteristics were evaluated in triplicate and the average values were used in data analysis.

Mechanical Properties

The tensile properties were determined using an Instron universal testing machine (model 5569, Instron Corp., USA) with a crosshead speed of 500 mm/min., and 1-kN load cell. The specimens were stamp-cut from a 2-mm-thick compression-moulded sheet. The dimension of the test specimens used was type I according to ISO 37 [22]. The specimens were symmetrically placed at the grips of the testing machine to achieve uniform tension distribution over the cross section. The tensile strength was determined from stress at rupture while the modulus at 100% strain was evaluated from the tensile stress at 100% elongation. The elongation at break was also determined.

The tear strength, a measure of the resistance of a material to tear force, was measured with a Lloyd instrument (model LS500-9674, Lloyd Instruments Ltd., UK) according to ISO 34-1 [23] using type-B die. Nicked-tab-end specimens were cut from a 2-mm-thick compression-moulded vulcanised sheet. The tear strength was tested at a crosshead speed of 500 mm/min. using a 1-kN load cell.

The sample hardness was determined using a Shore A durometer (model HPE-A, Bareiss, Germany) in accordance with ASTM D2240-05 [24] It was determined at three different positions on the specimens (about 6-mm thick) and the median value was indicated.

An abrasion test was carried out according to DIN 53516 [25] using an abrasion tester (model AB 6252, Bareiss, Germany). The abrasion resistance of a sample was expressed as volume loss when a cylindrically shaped specimen of 6-mm thickness is abraded for an abrasion distance of 40 m with emery paper (60 grit) at a constant force of 10 N.

Rebound resilience was determined according to DIN 53512 [26] using a rebound tester (model Rebound Check-Pendolo Shob, Gibitre Instruments S.r.l., Italy) and a cylindrically shaped specimen of 13-mm thickness. Rebound resilience was calculated as follows: Percentage resilience $= (1 - \cos\alpha) \times 100$ where α is the maximum rebound angle [26].

Each mechanical property test was repeated five times and an average value was used in the data analysis.

Scanning Electron Microscopy (SEM)

The phase morphology of the NR vulcanisates filled with different types of filler was examined using a scanning electron microscope (model JSM-5410LV, JEOL Ltd., Japan). Samples were cryogenically fractured in liquid nitrogen and then coated with a thin gold layer to prevent electrostatic charge during examination.

RESULTS AND DISCUSSION

Cure Characteristics

The cure characteristics of all NR compounds are shown in Table 5. The DRB-filled compound, with the highest torque difference, was comparable to the $CaCO_3$-filled compound. In general, the torque difference is an indicator of cross-link density of the vulcanisates [4]. Thus, a high torque difference of the DRB-filled compound indicated its high cross-link density.

It can be observed that the addition of fillers decreased the cure time of the rubber compounds. At a similar loading level of 50 phr, the DRB-filled compound exhibited the shortest cure time while compounds with clay and $CaCO_3$ showed comparable t_{90}. The scorch time of the filled NR compounds showed a trend similar to the cure time.

Table 5. Cure characteristics of different NR compounds

Compound Type	Torque difference, MH-ML (dN m)	Scorch time, t_{s2} (min.)	Cure time, t_{90} (min.)
Control	12.59	3.30	8.08
DRB-filled	18.78	2.16	4.18
Clay-filled	16.25	3.23	6.35
$CaCO_3$-filled	18.36	3.03	6.31

Mechanical Properties

Tensile Properties

The tensile properties of the different NR vulcanisates aer shown in Table 6. The tensile strength of all filled vulcanisates decreased because of the inability of the fillers to support stress transferred from the rubber matrix [14]. At a similar filler loading, clay gave the highest tensile strength, which corresponds to its highest surface area, as shown in Table 2, followed by DRB. Sae-

Oui et al. [27] reported that the surface area is the most important factor controlling the tensile strength.

The effect of fillers on the modulus, an indication of material stiffness, at 100% elongation is also shown in Table 6. The vulcanisate with DRB showed highest modulus and was comparable to that with clay, while clay gave elongation at break comparable to that of the control, with DRB giving a lower value.

Table 6. Tensile properties of NR vulcanisates

Vulcanisate	Tensile properties		
	100% Modulus (MPa)	Tensile strength (MPa)	Elongation at break (%)
Control	0.97 ± 0.07	20.09 ± 1.29	559 ± 71
DRB-filled	1.85 ± 0.14	7.09 ± 0.13	500 ± 28
Clay-filled	1.80 ± 0.53	17.50 ± 0.53	594 ± 1
$CaCO_3$-filled	1.44 ± 0.15	4.39 ± 0.72	360 ± 29

Tear Strength

A negative effect on the tear strength (Figure 1) was observed when the NR vulcanisate was filled with different fillers. Similar to tensile strength, the addition of the fillers somewhat reduced the tear strength of the vulcanisate. The results indicate that the tear strength seemed to be affected by the surface area of the fillers (Table 2) and also probably by a low rubber-filler interaction. At similar filler loading, clay gave the highest tear strength, followed by DRB.

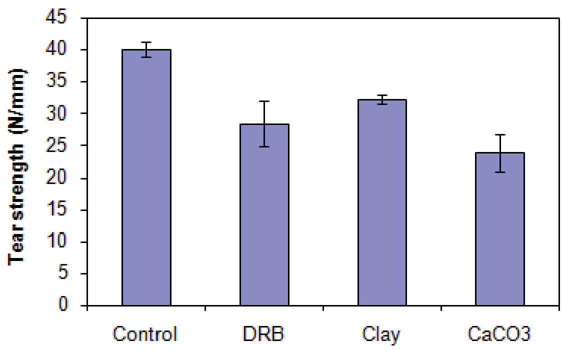


Figure 1. Tear strength of NR vulcanisates with different fillers (DRB, clay and $CaCO_3$)

Hardness and Rebound Resilience

The hardness values of the NR vulcanisates are shown in Figure 2. Compared to control, the filled vulcanisates exhibited higher hardness values, with DRB giving the highest value, followed by $CaCO_3$. The hardness values are seen to correspond to the torque difference values (Table 5), which indicates that the improvement in hardness of the filled vulcanisates was caused by an increase in the cross-link density of the vulcanisates. On the other hand, it can be seen (Figure 3) that the resilience of the vulcanisates, i.e. the elasticity of the rubber chain, slightly decreased with the addition of a filler.

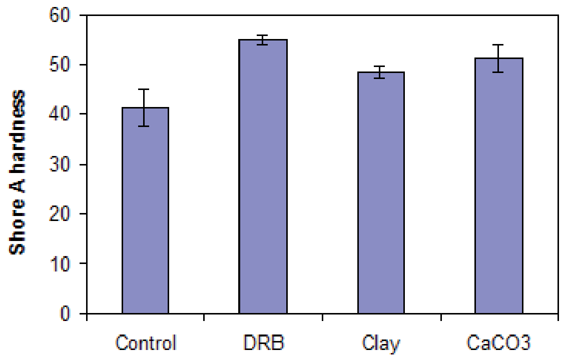


Figure 2. Hardness of NR vulcanisates with different fillers (DRB, clay and $CaCO_3$)

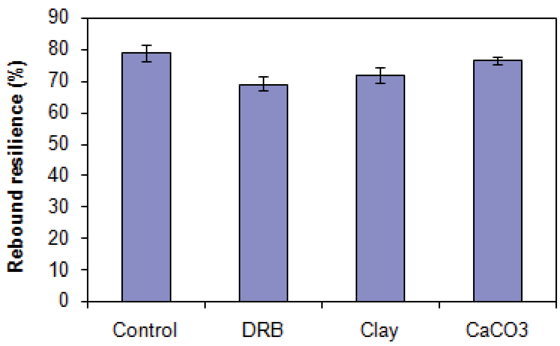


Figure 3. Rebound resilience of NR vulcanisates with different fillers (DRB, clay and $CaCO_3$)

Abrasion

The abrasion resistance of a solid body is defined as its ability to withstand the progressive removal of material from its surface as a result of the mechanical action of rubbing and scraping or of an erosive action [14]. The abrasion resistance of the NR vulcanisates, expressed as volume loss, is shown in Table 7; a higher volume loss means a lower abrasion resistance. The DRB-filled vulcanisate exhibited highest abrasion resistance and the control the lowest. The highest abrasion resistance of the DRB-filled vulcanisate corresponded to its highest hardness and cross-link density (Figure 2 and Table 5) while the control, with lowest hardness and cross-link density, also showed lowest abrasion resistance. Rattanasom and Chaikumpollert [28] also reported that the abrasion resistance of vulcanisates was contingent upon their hardness and cross-link density.

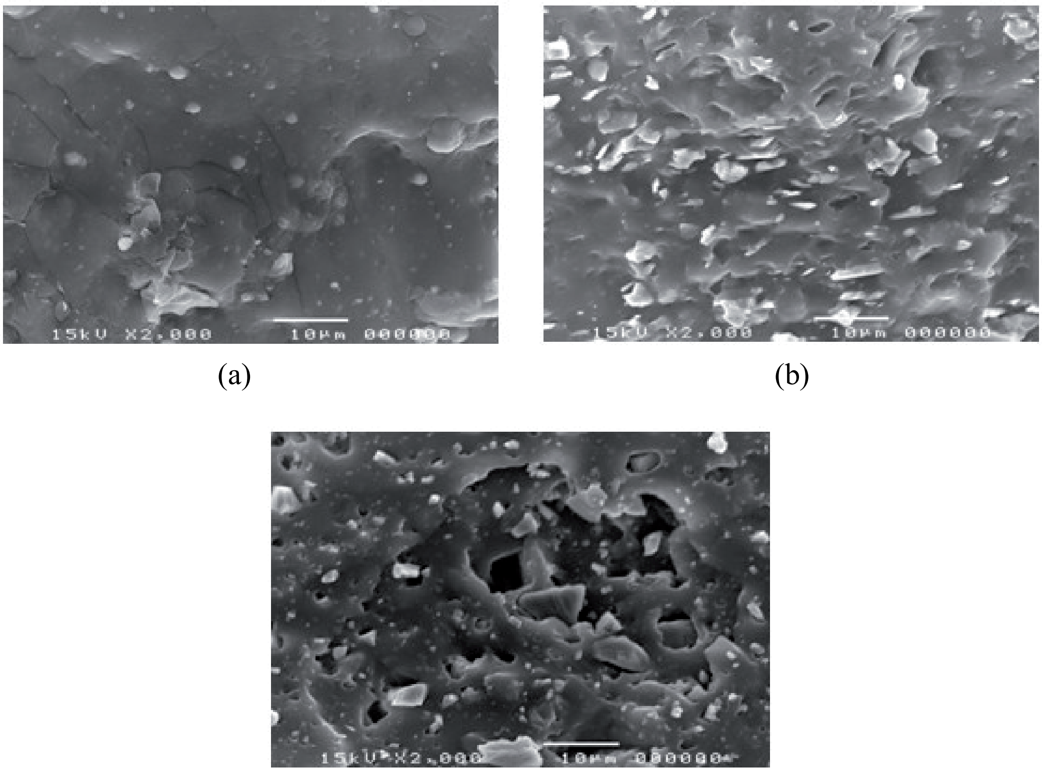
Morphology Study

Figure 4 shows scanning electron micrographs of the fractured surfaces of the NR vulcanisates blended with different types of fillers at 50-phr loading. Figure 4(a) shows an agglomeration of filler particles in the DRB-filled NR vulcanisate. This is expected because the

Table 7. Volume loss of NR vulcanisates

Vulcanisate	Volume loss (mm^3)
Control	66.27 ± 12.08
DRB-filled	38.05 ± 5.35
Clay-filled	51.08 ± 4.95
$CaCO_3$-filled	50.92 ± 5.07

interfacial interaction was weak due to the hydrophobic character of NR and the hydrophilic character of DRB. In the case of $CaCO_3$ filler, the filled NR vulcanisate showed several holes on the fractured surface on which the $CaCO_3$ particles were left (Figure 4(c)), which suggests that the interfacial interaction between the filler particles and the rubber was weak, resulting in the deterioration of the vulcanisate. Arayapranee and Rempel [14] also reported that the use of $CaCO_3$ as filler in NR/EPDM blends gave rise to many holes on the fractured surface due to a weak interfacial interaction. They found deterioration of the blend properties such as tensile and tear strength when compared with unfilled NR/EPDM blends. Figure 4(b) reveals that the clay particles were well dispersed without agglomeration within the rubber matrix. The dispersion of clay was better than that of the other fillers, thus contributing to a greater tensile and tear strength of the clay-filled vulcanisate.



(a) (b)

(c)

Figure 4. Scanning electron micrographs of fractured surfaces of NR vulcanisates filled with (a) DRB, (b) Clay and (c) $CaCO_3$

CONCLUSIONS

DRB was found to be comparable to clay and $CaCO_3$ with respect to its properties as a filler of an NR vulcanisate. Thus, it should have a high potential for being utilised as such, especially when coupled with the fact that it is cheap as well as renewable.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Thailand Research Fund (TRF) for a research grant (RDG5150042).

REFERENCES

1. A. K. Bagghi and B. G. Sharma, "Reinforcement and physical properties of filled rubber system", *Indian J. Technol.*, **1981**, *19,* 368-372.
2. C. S. L. Baker, "Properties of natural rubber with some highly reinforcing carbon blacks", *NR Technol.*, **1978**, *8,* 24-32.
3. D. C. Edwards, "Review polymer-filler interactions in rubber reinforcement", *J. Mater. Sci.*, **1990**, *25,* 4175-4185.
4. N. Rattanasom and S. Prasertsri, "Relationship among mechanical properties, heat ageing resistance, cut growth behaviour and morphology in natural rubber: Partial replacement of clay with various types of carbon black at similar hardness level", *Polym. Test.*, **2009**, *28,* 270-276.
5. R. Rajasekar, K. Pal, G. Heinrich, A. Das and C. K. Das, "Development of nitrile butadiene rubber-nanoclay composites with epoxidized natural rubber as compatibilizer", *Mater. Design*, **2009**, *30*, 3839-3845.
6. A. Usuki, A. Tukigase and M. Kato, "Preparation and properties of EPDM-clay hybrids", *Polymer*, **2002**, *43*, 2185-2189.
7. F.-L. Jin and S.-J. Park, "Thermo-mechanical behaviors of butadiene rubber reinforced with nano-sized calcium carbonate", *Mater. Sci. Eng. A*, **2008**, *478*, 406-408.
8. M. Mondragon, E. M. Hernandez, J. L. Rivera-Armenta and F. J. Rodriguez-Gonzalez, "Injection molded thermoplastic starch/natural rubber/clay nanocomposites: Morphology and mechanical properties", *Carbohyd. Polym.*, **2009**, *77*, 80-86.
9. N. Sharma, L. P. Chang, Y. L. Chu, H. Ismail, U. S. Ishiaku and Z. A. Mohd Ishak, "A study on the effect of pro-oxidant on the thermo-oxidative degradation behaviour of sago starch filled polyethylene", *Polym. Degrad. Stabil.*, **2001**, *71*, 381-393.
10. Y.-P. Wu, M.-Q. Ji, Q. Qi, Y.-Q. Wang and L.-Q. Zhang, "Preparation, structure, and properties of starch/rubber composites prepared by co-coagulating rubber latex and starch paste", *Macromol. Rapid Comm.*, **2004**, *25*, 565-570.
11. G. H. Yew, A. M. Mohd Yusof, Z. A. Mohd Ishak and U. S. Ishiaku, "Water absorption and enzymatic degradation of poly(lactic acid)/rice starch composites", *Polym. Degrad. Stabil.*, **2005**, *90*, 488-500.
12. M. Razavi-Nouri, F. Jafarzadeh-Dogouri, A. Oromiehie and A. E. Langroudi, "Mechanical properties and water absorption behaviour of chopped rice husk filled polypropylene composites", *Iran Polym. J.*, **2006**, *15*, 757-766.

13. A. I. Khalf and A. A. Ward, "Use of rice husks as potential filler in styrene butadiene rubber/linear low density polyethylene blends in the presence of maleic anhydride", *Mater. Design*, **2010**, *31*, 2414-2421.
14. W. Arayapranee and G. L. Rempel, "A comparison of the properties of rice husk ash, silica, and calcium carbonate filled 75:25 NR/EPDM blends", *J. Appl. Polym. Sci.*, **2008**, *110*, 1165-1174.
15. L. Chotirat, K. Chaochanchaikul and N. Sombatsompop, "On adhesion mechanisms and interfacial strength in acrylonitrile-butadiene-styrene/wood sawdust composites", *Int. J. Adhes. Adhes.*, **2007**, *27*, 669-678.
16. H. Ismail, Salmah and M. Nasir, "Dynamic vulcanization of rubberwood-filled polypropylene/natural rubber blends", *Polym. Test.*, **2001**, *20*, 819-823.
17. C. Pattamaprom, K. Bandidchutikun, S. Sotananan and S. Phrommedetch, "The effect of rubber modification on physical properties of rice husk ash composites", *Thammasat Int. J. Sc. Technol.*, **2008**, *13,* 36-43.
18. A. K. Mohanty, Q. Wu and S. Selke, "Novel 'green' materials from soy meal and natural rubber blends", *US Patent 2006/0041036 A1* (**2005**).
19. J. Wiboonsirikul, Y. Kimura, M. Kadota, H. Morita, T. Tsuno and S. Adachi, "Properties of extracts from defatted rice bran by its subcritical water treatment", *J. Agric. Food Chem.*, **2007**, *55*, 8759-8765.
20. ISO 9277, "Determination of the Specific Surface Area of Solids by Gas Adsorption-BET", The British Standards Institution, **2010**.
21. ISO 6502, "Rubber-Guide to the Use of Curemeters", The British Standards Institution, **1999**.
22. ISO 37, "Rubber, Vulcanized or Thermoplastic Determination of Tensile Stress-Strain Properties", The British Standards Institution, **2005**.
23. ISO 34-1, "Rubber, Vulcanized or Thermoplastic Determination of Tear Strength. Part 1: Trouser, Angle and Crescent Test Pieces", The British Standards Institution, **2004**.
24. ASTM D2240-05, "Standard Test Method for Rubber Property-Durometer Hardness.", ASTM International, **2010**.
25. DIN 53516, "Testing of Rubber and Elastomer—Determination of Abrasion Resistance", The German Institute for Standardization, **1977**.
26. DIN 53512, "Testing of Rubber and Elastomer—Determining the Rebound Resilience of Rubber Using the Schob Pendulum", The German Institute for Standardization, **1988**.
27. P. Sai-Oui, C. Rakdee and P. Thanmathorn, "Use of rice husk ash as filler in natural rubber vulcanizates: In comparison with other commercial fillers", *J. Appl. Polym. Sci.*, **2002**, *83,* 2485-2493.
28. N. Rattanasom and O. Chaikumpollert, "Crack growth and abrasion resistance of carbon black-filled purified natural rubber vulcanizates", *J. Appl. Polym. Sci.*, **2003**, *90,* 1793-1796.

Maejo Int. J. Sci. Technol. **2012**, *6*(02), 259-271

Maejo International Journal of Science and Technology

ISSN 1905-7873

Available online at www.mijst.mju.ac.th

Full Paper

Resistant starch content, in vitro starch digestibility and physico-chemical properties of flour and starch from Thai bananas

Nednapis Vatanasuchart*, Boonma Niyomwit and Karuna Wongkrajang

Department of Nutrition and Health, Institute of Food Research and Product Development, Kasetsart University, Bangkok 10900, Thailand

* Corresponding author, e-mail: ifrnpv@ku.ac.th

Received: 20 May 2011 / Accepted: 29 June 2012 / Published: 17 July 2012

Abstract: Flour and starch were prepared from six Thai banana cultivars: Kluai Hom, Kluai Khai, Kluai Lebmuenang, Kluai Namwa, Kluai Hakmuk and Kluai Hin, and their resistant starch (RS), in vitro starch digestibility and physico-chemical properties were determined. The RS content of the flour is 52.2-68.1%, with flour from Kluai Hin containing the highest amount of RS, followed by that from Kluai Hakmuk. The starch has a higher RS content (70.1-79.2%), the highest value coming from Kluai Hakmuk starch, followed by Kluai Hom starch. A significant linear relationship between apparent amylose and RS was observed. Interestingly, most of the flour showed a slower rate of in vitro starch digestibility than that of the starch, with Kluai Hin flour exhibiting the slowest rate, followed by Kluai Namwa. Rapid viscosity analysis showed significantly higher peak viscosity of the starch than the flour, the highest final and setback viscosity being obtained from Kluai Hin starch. Differential scanning calorimetry showed an endothermic transition enthalpy over a range of 17.4 J/g for Kluai Lebmuenang starch to 18.6 J/g for Kluai Hin starch. X-ray diffractograms of the starches exhibited a typical B-pattern with Kluai Hin showing the highest degree of relative crystallinity (31.3%) with a sharp peak at 5.5°. The overall results seemed to indicate an effect of the BB genotype on the resistance of banana starch granules to enzymatic digestion due to amylose molecules and the crystallinity of amylopectin.

Keywords: Thai bananas, green bananas, banana flour, banana starch, resistant starch, starch digestibility

---

INTRODUCTION

Bananas (*Musa* sp.) are one of the most important tropical fruits consumed worldwide by people of all age groups. The nutritional and functional properties of bananas are known to provide good health. Nutritionally, bananas contain available carbohydrates which provide energy, vitamins B and C, and significant amounts of potassium and magnesium [1]. A substantial percentage of starch in bananas consists of resistant starch (RS), which has the potential to provide significant health benefits akin to those derived from dietary fibre [2]. Due to a high solid content of 40-70% [3], bananas can be processed into flour and starch suitable for making processed health food products.

At present, a healthy choice of functional food products is of increasing interest to consumers. With properties similar to soluble and insoluble dietary fibre in the gastrointestinal tract, RS plays a major role in the health food industry [4-5]. Showing some resistance to human digestive enzymes, the slow release of glucose from RS results in reduced energy intake by the intestinal cells, which is evident from a low glycemic index of the non-digested starch [6]. This can help improve glucose regulation in diabetes and facilitate weight control for the obese [7]. The non-digested starch in the large intestine is fermented by colonic microflora, producing short-chain fatty acids that encourage the growth of beneficial bacteria [1]. This may lead to healthier colon cells and help prevent the development of colon cancer [4]. In addition, a diet high in RS can reduce blood cholesterol and triglyceride levels due to higher excretion rates of cholesterol and bile acids [7]. Overall, increasing the RS content in the diet has the potential to provide several significant health benefits and an added value to food products.

RS is defined as the sum of starch and products of starch degradation not absorbed in the small intestine of a healthy individual [8-9]. There are four types of RS: type I represents physically inaccessible starch which is locked in the plant cell walls of some foodstuffs such as partially milled grains, seeds and legumes. Type II is characterised by native granular starch found in foods containing uncooked starch such as bananas, raw potatoes and beans. The RS content in reference banana flour samples, determined by three laboratories, averages 52.1% (dry matter), while lentil flour has 8.2% RS [10]. A study of Vatanasuchart et al. [11] on eleven banana cultivars grown in Thailand also shows that the RS content observed in the common cultivars ranges between 52.2-61.4% and values for indigenous cultivars are between 50.7-68.1%. Type-III RS is made up of retrograded starch or crystalline non-granular starch such as that found in cooked potatoes, bread crust, cornflakes and retrograded high-amylose maize starch. Type-IV RS refers to specific chemically and thermally modified or repolymerised starch [8-9].

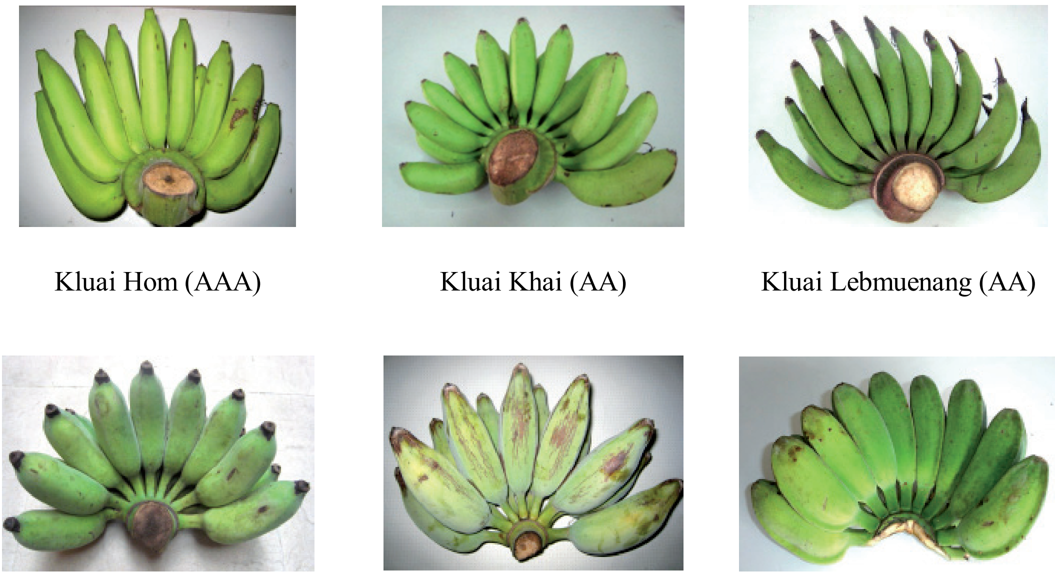
The properties of banana starch and flour are important for their utilisation in industrial food products. Several reports have suggested that banana starch contains a high level of amylose which is associated with high retrogradation [1, 11-12]. From flour of the common and indigenous banana cultivars, Vatanasuchart et al. [10] reported a high content of apparent amylose (24.7-37.1% and 29.9-35.9% respectively). The findings of Nimsung et al. [3], however, indicated a lower amylose content in banana starch isolated from Kluai Khai and Kluai Hom cultivars. A study of Tongdang and Saasagul [13] showed two stages of swelling of starch from Kluai Hin and Kluai Namwa cultivars as a result of a higher proportion of amylose in relation to amylopectin, while the gelatinisation enthalpy of Kluai Hin (20.16 J/g) was higher than Kluai Namwa (17.43 J/g).

As starchy foods are a main source of energy in the diet of Thai people, a healthier choice of starchy foods that still provide beneficial functions for sustaining good health should be encouraged. Therefore, it is relevant to acquire new knowledge about the health benefits of different banana cultivars grown in Thailand and to hypothesise that both starch and flour from bananas are a good source of RS which is good for health. Our previous study [11] on RS and amylose in banana flour concerned only a chemical aspect, so more data on the functional properties of both banana starch and flour should be of use. In this study, the contents of total starch, digestible starch and RS (non-digestible starch) in six banana cultivars were determined. The rates of in vitro starch digestibility and the physico-chemical properties of banana starch and flour samples were also compared. X-ray diffraction patterns of banana starch samples were examined.

MATERIALS AND METHODS

Sample Preparation

Edible green (unripe) bananas (six cultivars), aged 90-120 days, were collected from Pakchong Research Station, Kasetsart University at Nakhon Ratchasima, and also from local markets. Different species of diploid and triploid genome groups were classified into four common cultivars: Kluai Hom (AAA), Kluai Khai (AA), Kluai Lebmuenang (AA) and Kluai Namwa (ABB), and two indigenous cultivars: Kluai Hakmuk (ABB) and Kluai Hin (BBB) [14] (Figure 1). In processing into flour, the bananas were peeled and sliced into 1-mm-thick pieces, spread evenly on a stainless steel tray, dried in a hot-air oven at 50°C for 8 hr, and then milled and passed through a 100-mesh sieve. Banana starch was prepared according to a water-alkaline extraction process adopted by Zhang et al. [15], as presented in Scheme 1.



Kluai Hom (AAA) | Kluai Khai (AA) | Kluai Lebmuenang (AA)

Kluai Namwa (ABB) | Kluai Hakmuk (ABB) | Kluai Hin (BBB)

Figure 1. Different Thai banana cultivars. (AA represents two sets of chromosomes and AAA, ABB or BBB represent three sets of chromosomes inherited from their parents.)

Macerate peeled and sliced banana (500 g) with 0.05N NaOH (2 L)
at high speed for 1 min.

↓

Screen homogenate through doubled sheets of muslin,
collect filtrate and let stand for 2 hr

↓

Decant supernatant of dark material and add water (8 L) to sediment,
stir thoroughly and let stand for 2 hr

↓

Again decant supernatant and screen residue through 120-μm nylon

↓

Add deionised water (8 L) to filtrate, stir thoroughly and let stand overnight

↓

Decant supernatant and collect starch portion, add deionised water (8 L),
stir thoroughly and let stand for 2 hr

↓

Decant clear supernatant and transfer starch portion to beaker, then
add deionised water (1 L) and stir thoroughly

↓

Transfer starch suspension to centrifuge tubes

↓

Centrifuge at 8000 rpm for 15 min. and collect sediment

↓

Dry using hot air at 40°C for 12 hr and sieve banana starch through 100-mesh sieve

Scheme 1. Process for isolation of banana starch

Reagents and Chemicals

The hydrolytic enzymes of pancreatic α-amylase (Sigma, no.A3176; 23 IU/mg), pepsin (Merck, no.7190; 2000 FIP-U/g) and amyloglucosidase (Boehringer, no.102857) as well as a glucose oxidase-peroxidase kit (Sigma, no.G3660) were used for RS determination and in vitro starch digestibility test in this study. Amylose and amylopectin used as standards were purchased from Sigma. Other chemical agents were of analytical grade from Merck.

Determination of RS, Total Starch and Amylose

The RS content of banana samples was determined by a direct method of Goni et al. [10]. A ground sample of flour or starch (100 mg) was incubated with a solution containing 20 mg pepsin at 40°C for 60 min. to remove protein. A tris-maleate solution containing 40 mg pancreatic α-amylase was then added and the mixture incubated at 37°C for 16 hr to hydrolyse digestible starch. The hydrolysate was centrifuged and the residue was solubilised with 4M KOH and incubated with amyloglucosidase (80 μL) at 60°C for 45 min. to hydrolyse RS. The glucose content was measured using a glucose oxidase-peroxidase kit. The RS content, comprising fractions of types I and II, was calculated as mg of glucose x 0.9.

The total starch content was determined according to a modified method of Goni et al. [16]. A 50-mg ground sample of flour or starch was dispersed in 2M KOH (6 mL) and the mixture incubated for 30 min. at room temperature. The solubilised starch was then hydrolysed by adding amyloglucosidase (60 μL) and incubating at 60°C for 45 min. in a shaking water bath. After centrifugation (15 min., 4500*g*), the glucose content in the supernatant was measured using the glucose oxidase-peroxidase kit, and the total starch content was calculated as mg of glucose x 0.9. Digestible starch content was calculated as the difference between total starch and RS or indigestible starch, expressed as per cent of the sample dry weight.

The amylose content of banana flour and starch samples was determined by a colorimetric AACC method [17]. Briefly, a 100-mg sample was gelatinised in the presence of 95% ethanol (1 mL) and 1N NaOH (9 mL) to liberate amylose molecules. Iodine solution (2 mL) was added to form an amylose-iodine complex and absorbance was read at 620 nm. The amylose content was calculated by means of a standard curve and expressed as per cent of sample dry weight.

In vitro Starch Digestibility

Following the method of Goni et al. [16], a ground sample (50 mg) of banana flour or starch was incubated with a solution containing pepsin (20 mg) at 40°C for 60 min. to remove protein, and the volume of the mixture was made up to 25 mL with tris-maleate buffer. Five mL of tris-maleate solution containing α-amylase (3.3 IU) were then added and the mixture incubated at 37°C to hydrolyse digestible starch. An aliquot sample (1 mL) was taken every 30 min. during a 3-hr period and heated at 100°C for 5 min. to inactivate the enzyme. Then amyloglucosidase (60 μL) was added to each of the aliquote samples to hydrolyse the remaining starch at 60°C for 45 min. After centrifugation (15 min., 4500*g*), the glucose content in the supernatant was measured using the glucose oxidase-peroxidase kit, and the digestible starch was calculated as mg of glucose x 0.9. The digestibility was expressed in terms of glucose release per 100 g of sample hydrolyzed at different times (0, 30, 60, 120, 160 and 180 min).

Pasting Properties, Thermal Properties and X-ray Diffraction Measurement

The pasting properties of 10% starch and flour suspensions in distilled water were determined by using a rapid viscosity analyser (RVA 4D, Newport Scientific, Australia). The determination was started at 50°C; the suspension was then heated to 95°C, maintained for 2.5 min., cooled to 50°C and held for 2 min. The total running time for each sample was 13 min [18].

Thermal properties were determined according to a method adapted from Vatanasuchart et al. [19] using a differential scanning calorimeter (Pyris I, Perkin-Elmer) equipped with a cooling system. Each starch sample (3 mg) was weighed in an aluminum pan and deionised water was added to obtain a 30% starch suspension. The cover was put on and hermetically sealed. Each sample pan was placed in the calorimeter and heated from 0°C to 120°C at 10°C/min. An empty pan was used as a reference and the instrument was calibrated using indium control. Endothermal curves exhibiting onset, peak and end temperatures and melting enthalpy (J/g of the sample weight on dry basis) of duplicate samples were recorded.

Wide-angle X-ray diffraction patterns of starch samples were examined using a X-ray diffractometer (JDX 3530, JEOL Ltd.) operated at 30 kV and 30 mA and generating monochromatic Cu-K_α radiation of 1.542 Å. Diffractograms were obtained from 4° to 30° (2θ) at a scanning speed of 4°/min. The degree of relative crystallinity was calculated from the ratio of diffraction peak area to total diffraction area [20].

RESULTS AND DISCUSSION

Starch Content

The total, digestible, and resistant starch contents of flour and starch from the six banana cultivars are shown in Table 1. The RS content of the flour ranges between 52.2-68.1% and that of the starch, 70.1-79.2%. Kluai Hin flour has the highest RS content, followed by Kluai Hakmuk. Most of the starch samples have significantly higher RS content than that of the flour. Study by Englyst et al. [5] and Goni et al. [10] on RS in banana flour showed values of 51.3-53.1%. Also, work by Tribess et al. [21] showed a high RS content in banana flour (40.9-58.5%), similar to the findings by Faisant et al. (47.3-57.2%) [2], whereas Rodríguez-Ambriz et al. [22] found a lower RS content (30.4%). However, the overall results from the present study indicate that the indigenous Kluai Hakmuk (ABB) and Kluai Hin (BBB) cultivars in the BB genome group are rich in RS when compared to the common cultivars.

Table 1. Total starch, digestible starch, RS and amylose contents (g/100g dry weight) of different banana cultivars

Sample	Total starch	Digestible starch	RS	Amylose
Flour:				
Kluai Hom (AAA)	91.0 ± 3.1 ^{b}	33.3 ± 1.9 ^{a}	57.7 ± 1.1 ^{de}	25.0 ± 0.3 ^{c}
Kluai Khai (AA)	80.5 ± 0.3 ^{c}	28.2 ± 4.5 ^{ab}	52.2 ± 4.1 ^{e}	24.7 ± 2.5 ^{c}
Kluai Lebmuenang (AA)	72.1 ± 3.4 ^{d}	15.1 ± 3.2 ^{d}	57.0 ± 0.2 ^{de}	31.2 ± 0.8 ^{b}
Kluai Namwa (ABB)	79.7 ± 1.1 ^{c}	23.0 ± 4.8 ^{bc}	56.6 ± 5.8 ^{de}	25.8 ± 2.7 ^{bc}
Kluai Hakmuk (ABB)	72.3 ± 1.8 ^{d}	10.9±0.5 ^{def}	61.4 ± 2.3 ^{d}	27.3 ± 2.8 ^{bc}
Kluai Hin (BBB)	72.7 ± 1.4 ^{d}	4.6 ± 0.6 ^{f}	68.1 ± 2.0 ^{c}	29.9 ± 0.6 ^{bc}
Starch:				
Kluai Hom (AAA)	82.7 ± 1.0 ^{c}	6.6 ± 0.3 ^{ef}	76.1 ± 0.7 ^{ab}	38.6 ± 0.5 ^{a}
Kluai Khai (AA)	96.0 ± 2.1 ^{a}	25.5 ± 3.3 ^{b}	70.5 ± 0.0 ^{bc}	40.9 ± 0.2 ^{a}
Kluai Lebmuenang (AA)	98.0 ± 1.7 ^{a}	22.5 ± 1.4 ^{bc}	75.5 ± 3.2 ^{ab}	42.7 ± 0.5 ^{a}
Kluai Namwa (ABB)	88.9 ± 1.0 ^{b}	13.3 ± 1.6 ^{de}	75.6 ± 0.6 ^{ab}	43.8 ± 1.1 ^{a}
Kluai Hakmuk (ABB)	87.4 ± 1.3 ^{b}	8.1 ± 2.6 ^{ef}	79.2 ± 3.9 ^{a}	39.3 ± 1.6 ^{a}
Kluai Hin (BBB)	87.0 ± 0.4 ^{b}	16.9 ± 6.7 ^{cd}	70.1 ± 4.9 ^{bc}	40.3 ± 7.0 ^{a}

Notes: 1) Values are means of duplicate analysis. In a column, means not sharing a common letter are significantly different at $P < 0.05$ by ANOVA and DMRT.

2) AAA, AA, etc. are genotypes of banana cultivars.

In Vitro Starch Digestibility

The experimental values for amylose content of the banana flour and starch were 24.7-31.2% and 38.6-43.8% respectively (Table 1). When the apparent amylose content of both the banana flour and starch was compared to the RS content, a significant linear relationship was observed, with $R^2 = 0.76$ ($p<0.05$) (Figure 3). This indicates that the resistance to enzymatic digestion of banana starch granules comes from amylose molecules. A recent study on four Indian red lentil cultivars [23] also showed a positive correlation between amylose and slow-digestible starch contents, and a negative correlation of amylose with RS contents.

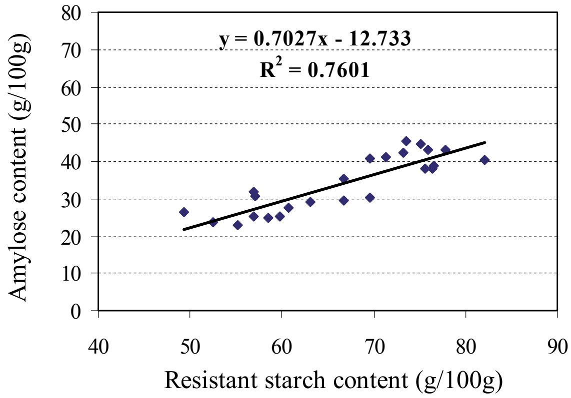


Figure 3. Linear correlation between RS content and apparent amylose content

The in vitro starch digestibility of banana flour and starch from the six cultivars, in comparison to cassava starch, is shown in Figure 4. The flour and starch of Kluai Hin were observed to show lowest digestibility during 30-180 min.of digestion. The present study also shows that the flour is more difficult to digest than the starch and that all banana flour and starch samples exhibit a lower starch digestibility compared to cassava starch.

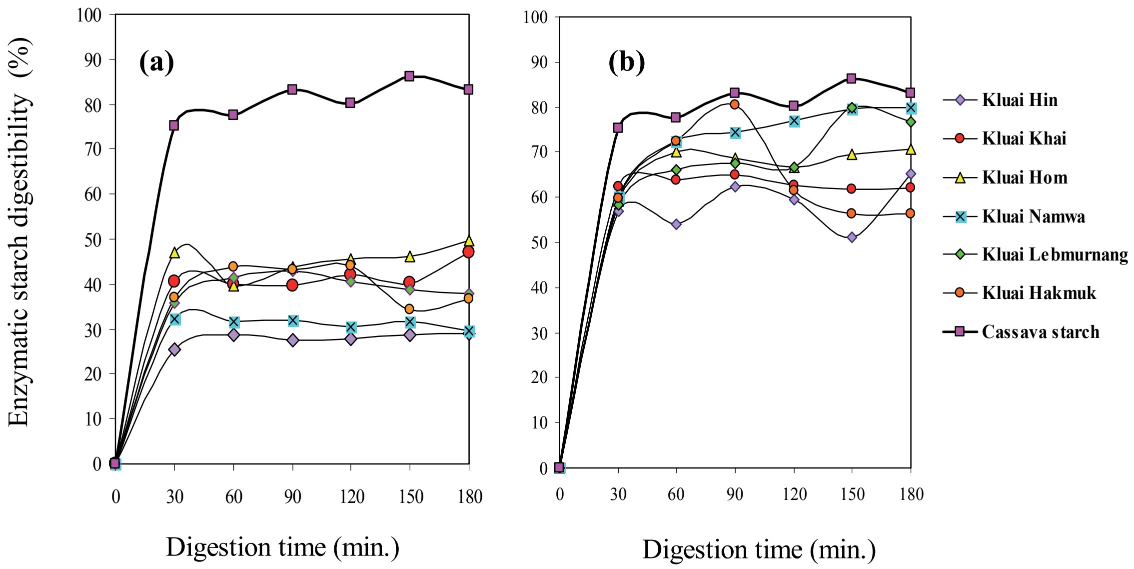


Figure 4. In vitro enzymatic starch digestibility of flour (a) and starch (b) from six banana cultivars as compared with cassava starch

Several studies have suggested that consumption of unripe bananas confers beneficial effects on human health, a fact often associated with their high content of RS, dietary fibre or non-starch polysaccharides, and nutritive and functional components [1, 2, 24]. In particular, pectin, lignin, cellulose and hemicellulose in banana fruit can promote digestive health [25-26]. The slow digestion rate found among the banana flour might well be explained by the complexation of pectin, cellulose and hemicellulose, which can help resist digestion better than the starch alone.

Pasting Properties

The pasting behaviours of the flour and starch from six banana cultivars are shown in Figure 5. Significantly higher peak viscosities of the starch samples (250.3-304.2 RVU) than those of the flour samples (153.4-213.5 RVU) were observed. Mostly, the banana flour showed a slight change in trough and breakdown viscosities whereas the starch gave greater final viscosities (278.3-356.2 RVU) than the flour (161.0-274.8 RVU), leading to lower setback viscosities of the flour. In particular, the flour from Kluai Hin showed lowest final and setback viscosities while its starch exhibited highest values. The high final viscosities of the starch from Kluai Hin, Kluai Namwa and Kluai Lebmuenang were in accordance with their high amylose content conducive to retrogradation. The pasting of the flour and starch occurred at 79.1-85.2°C and 79.2-83.2°C respectively. These temperatures were higher than for cassava starch (69.5°C), which has a lower amylose content [18]. As previously found, the structural stability of banana starch granules is influenced both by molecules of amylose and the degree of crystallinity in amylopectin [27-28].

Thermal Properties

The thermal characteristics of starch from six banana cultivars (Table 2) as determined by differential scanning calorimetry show that the endothermic gelatinisation enthalpy of Kluai Hin starch is highest and that the starch from Kluai Namwa (ABB) and Kluai Hin (BBB) with the BB genotype has higher thermal enthalpy and higher peak and end temperatures than the starch with AAA and AA genotypes. Also, a broader gelatinisation temperature range of starch from Kluai Namwa, Kluai Hakmuk and Kluai Hin can be observed. Likewise, the findings of Nimsung et al. [3] indicated a higher gelatinisation temperature for Kluai Namwa starch compared to Kluai Hom and Kluai Khai starches. In a similar study [13], the gelatinisation temperature range of Kluai Hin starch was found to be broader (69.8-79.8°C) than that of Kluai Namwa starch (70.9-77.2°C) with the former starch also having a higher gelatinisation enthalpy. The difference in gelatinisation temperature may be attributed to the difference in the amylose content as well as the difference in the size, form and distribution of starch granules [8, 9, 27, 29]. Shamai et al. [20] reported that starch gelatinisation temperature is related to the structural characteristics of crystallinity.

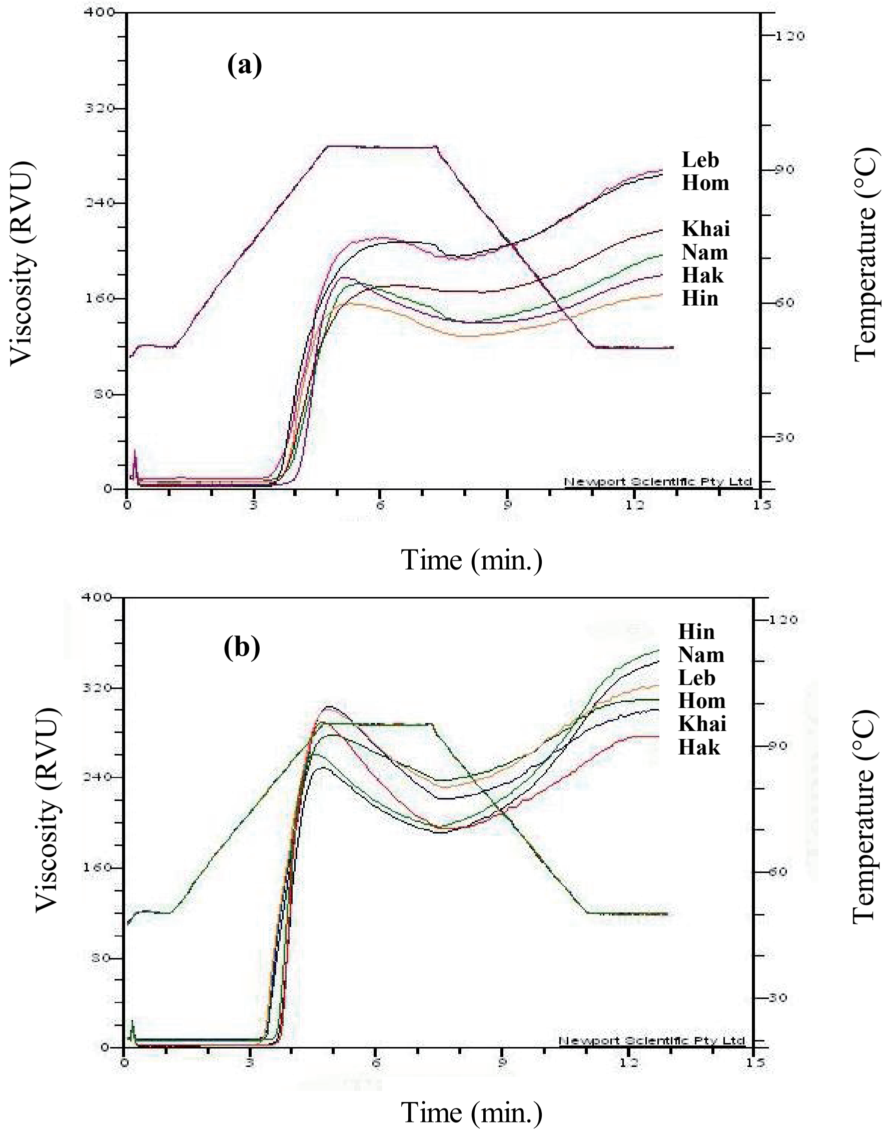


Figure 5. Viscosity curves of flour (a) and starch (b) of six banana cultivars: Kluai Hom (Hom), Kluai Khai (Khai), Kluai Lebmuenang (Leb), Kluai Namwa (Nam), Kluai Hakmuk (Hak) and Kluai Hin (Hin)

Table 2. Thermal properties of banana starch from different cultivars

Starch sample	Enthalpy (J/g)	Gelatinisation temperature (°C)		
		Onset	Peak	End
Kluai Hom (AAA)	18.1 ± 0.1	72.3 ± 0.3	74.7 ± 0.4	77.5 ± 0.5
Kluai Khai (AA)	17.5 ± 0.2	69.0 ± 0.2	72.2 ± 0.1	76.2 ± 0.3
Kluai Lebmuenang (AA)	17.4 ± 0.2	69.1 ± 0.2	72.1 ± 0.1	75.3 ± 0.2
Kluai Namwa (ABB)	18.3 ± 0.1	71.6 ± 0.1	75.8 ± 0.0	79.9 ± 0.1
Kluai Hakmuk (ABB)	17.7 ± 0.2	72.3 ± 0.1	77.1 ± 0.1	83.8 ± 0.1
Kluai Hin (BBB)	18.6 ± 0.1	72.0 ± 0.0	75.9 ± 0.1	79.6 ± 0.1

Note: Values are means of duplicate analysis.

X-ray Diffraction

X-ray diffractograms of banana starch from the six cultivars showed a typical B-pattern, with peaks at 2θ of about 5.5°, 15°, 17° and 23° (Figure 6). Kluai Hin starch gave the highest degree of relative crystallinity (31.3%) and exhibited a sharper peak at 5.5°. These findings indicate that starch from bananas with BB genotype seem to have a higher degree of relative crystallinity when compared to AA genotype banana starch.

Raw starch granules are semi-crystallites comprising amylose and amylopectin polymers. The degree and type of crystallinity present is dependent mainly on the structural characteristics of amylopectin and three types (A, B and C) of crystalline structure of starch granules have been distinguished [29]. According to several reports, banana starches can have A-type, B-type or a mixture of the two, depending on varietal source, growing conditions and other factors. However, B-type diffraction patterns for banana starch have often been reported [2,13,15].

In the present study, a B-type crystalline structure of banana starch granules was found, with a distinctive peak at 2θ of about 5.5°, which is considered to be a fingerprint for the B-type structure and is in agreement with the report by Shamai et al. [20]. Also, according to another study by Williamson et al. [30] on the enzymatic resistance of different polymorphs, the B-type structure is digested more slowly than the A-type by α-amylase. Thus, the B-type crystalline structure of banana starch observed in this study should be related to a high level of RS, which would consequently result in a high gelatinisation enthalpy, particularly for BB genotype bananas such as Kluai Hin and Kluai Namwa.

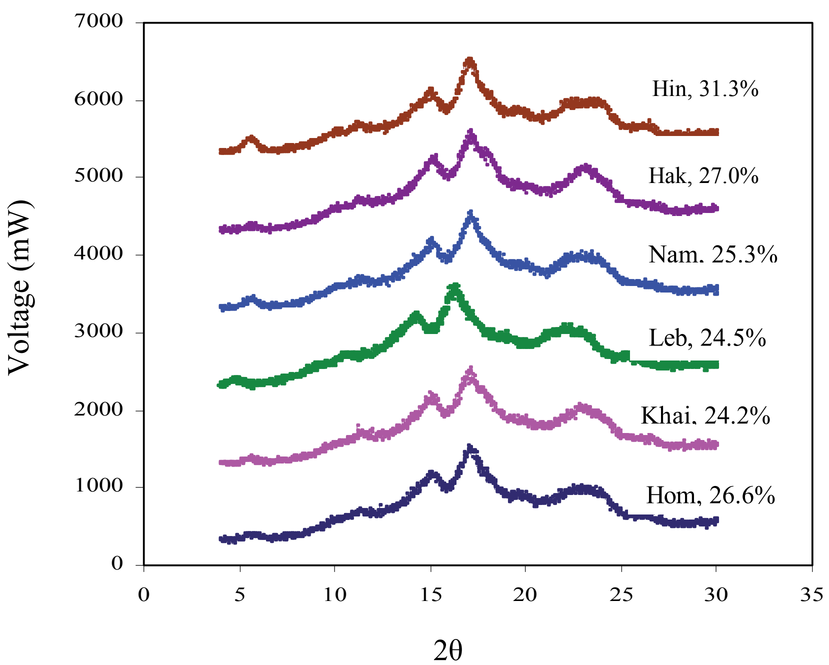


Figure 6. X-ray diffraction patterns of banana starch from different cultivars: Kluai Hom (Hom), Kluai Khai (Khai), Kluai Lebmuenang (Leb), Kluai Namwa (Nam), Kluai Hakmuk (Hak) and Kluai Hin (Hin), shown with degree of relative crystallinity in percentage.

CONCLUSIONS

The high content of RS found in both starch and flour samples of all six cultivars of Thai bananas indicates that they are a healthy choice for consumption. Their genotypes seem to influence the RS content and in vitro starch digestibility as well as physical and structural properties. The highest RS content was found in starch from Kluai Hakmuk (ABB genotype) and in flour from Kluai Hin (BBB genotype) (79.2% and 68.1% respectively). A significant linear relationship between apparent amylose and RS contents of the banana flour and starch samples was observed ($R^2 = 0.76$), indicating that the starch stability comes from amylose molecules. More importantly, Kluai Hin starch showed the slowest in vitro digestibility rate and highest thermal enthalpy. According to the X-ray diffraction patterns, this should result from a more confined structure of amylopectin crystallinity of Kluai Hin starch compared to other starches.

ACKNOWLEDGEMENTS

The authors are grateful to Kasetsart University and Thailand Research Fund (TRF) for their financial support, and also to Pakchong Research Station, Kasetsart University for the provision of banana samples.

REFERENCES

1. G. Aurore, B. Parfait and L. Fahrasmane, "Bananas, raw materials for making processed food products", *Trends Food Sci. Technol.*, **2009**, *20*, 78-91.
2. N. Faisant, A. Buléon, P. Colonna, C. Molis, S. Lartigue, J. P. Galmiche and M. Champ, "Digestion of raw banana starch in the small intestine of healthy humans: Structural features of resistant starch", *Br. J. Nutr.*, **1995**, *73*, 111-123.
3. P. Nimsung, M. Thongngam and O. Naivikul, "Compositions, morphological and thermal properties of green banana flour and starch", *Kasetsart J. (Nat. Sci.)*, **2007**, *41*, 324-330.
4. M. L. Ohr, "Nutraceuticals and functional foods: Fortifying with fiber", *Food Technol.*, **2004,** *58*, 71-75.
5. H. N. Englyst, S. M. Kingman and J. H. Cummings, "Classification and measurement of nutritionally important starch fractions", *Eur. J. Clin. Nutr.*, **1992**, *46*, 33-55.
6. M. Croghan, "Resistant starch–a better for your carbohydrate", Proceedings of Food Ingredients Asia Conference: Consumer Awareness on Healthy and Functional Ingredients, **2004**, Bangkok, Thailand, pp.1-17.
7. J. A. Higgins, D. R. Higbee, W. T. Donahoo, I. L. Brown, M. L. Bell and D. H. Bessesen, "Resistant starch consumption promotes lipid oxidation", *Nutr. Metab.*, **2004**, *1*, 8. (Available at: http://www.nutritionandmetabolism.com/content/pdf/1743-7075-1-8.pdf)
8. M. G. Sajilata, R. S. Singhal and P. R. Kulkarni, "Resistant starch─A review", *Compr. Rev. Food Sc. Food Safety*, **2006**, *5*, 1-17.
9. R. C. Eerlingen and J. A. Delcour, "Formation, analysis, structure and properties of type III enzyme resistant starch", *J. Cereal Sci.*, **1995**, *22*, 129-138.

10. I. Goni, L. García-Diz, E. Mañas and F. Saura-Calixto, "Analysis of resistant starch: A method for foods and food products", *Food Chem.*, **1996**, *56*, 445-449.
11. N. Vatanasuchart, B. Niyomwit and K. Wongkrajang, "Resistant starch contents and the in vitro starch digestibility of Thai starchy foods**",** *Kasetsart J. (Nat. Sci.)*, **2009**, *43*, 178-186.
12. R. V. da Mota, F. M. Lajolo, C. Ciacco and B. R. Cordenunsi, "Composition and functional properties of banana flour from different varieties", *Starch/Starke*, **2000**, *52*, 63-68.
13. T. Tongdang and T. Saasagul, "Some properties of banana starches extracted from two varieties", Proceedings of Starch Update 2005: The 3rd Conference on Starch Technology, **2005**, Bangkok, Thailand, pp. 199-200.
14. S. Achavachajchuan and C. Maksumpun, "Thai Bananas" (in Thai), Platapean Publishing, Bangkok, **2007**.
15. P. Zhang, R. L. Whistler, J. N. BeMiller and B. R. Hamaker, "Banana starch: Production, physicochemical properties and digestibility–A review", *Carbohydr. Polym.*, **2005**, *59*, 443-458.
16. I. Goni, A. Garcia-Alonso and F. Saura-Calixto, "A starch hydrolysis procedure to estimate glycemic index", *Nutr. Res.*, **1997**, *17*, 427-437.
17. American Association of Cereal Chemists, "AACC method 61-03.01: Amylose content of milled rice", in "Approved Methods of the American Association of Cereal Chemists" (Ed. K Cummings), AACC International, St. Paul MN, **2000,** pp.1-4.
18. N. Vatanasuchart, O. Naivikul, S. Charoenrein and K. Sriroth, "Effects of different UV irradiation on properties of cassava starch and biscuit expansion", *Kasetsart J. (Nat. Sci.)*, **2003**, *37*, 334-344.
19. N. Vatanasuchart, O. Naivikul, S. Charoenrein and K. Sriroth, "Molecular properties of cassava starch modified with different UV irradiation to enhance baking expansion", *Carbohydr. Polym.*, **2005**, *61*, 80-87.
20. K. Shamai, H. Bianco-Peled and E. Shimoni, "Polymorphism of resistant starch type III", *Carbohydr. Polym.*, **2003**, *54*, 363-369.
21. T. B. Tribess, J. P. Hernández-Uribe, M. G. C. Méndez-Montealvo, E. W. Menezes, L. A. Bello-Perez and C. C. Tadini, "Thermal properties and resistant starch content of green banana flour (*Musa cavendishii*) produced at different drying conditions", *LWT – Food Sci. Technol.*, **2009**, *42*, 1022-1025.
22. S. L. Rodríguez-Ambriz, J. J. Islas-Hernández, E. Agama-Acevedo, J. Tovar and L. A. Bello-Pérez, "Characterization of a fibre-rich powder prepared by liquefaction of unripe banana flour", *Food Chem.*, **2008**, *107*, 1515-1521.
23. M. Kaur, K. S. Sandhu and S. T. Lim, "Microstructure, physicochemical properties and in vitro digestibility of starches from different Indian lentil (*Lens culinaris*) cultivars", *Carbohydr. Polym.*, **2010**, *79*, 349-355.
24. K. Kanazawa and H. Sakakibara, "High content of dopamine, a strong antioxidant, in Cavendish banana**",** *J. Agric. Food Chem.*, **2000**, *48*, 844-848.
25. K. Kayisu, L. F. Hood and P. J. Vansoest, "Characterization of starch and fiber of banana fruit", *J. Food Sci.*, **1981**, *46*, 1885-1890.

26. J. L. Buttriss and C. S. Stokes, "Dietary fibre and health: An overview", *Nutr. Bull.*, **2008**, *33*, 186-200.
27. D. J. Gallant, B. Bouchet and P. M. Baldwin, "Microscopy of starch: evidence of a new level of granule organization", *Carbohydr. Polym.*, **1997,** *32*, 177-191.
28. J. Blazek and L. Copeland, "Pasting and swelling properties of wheat flour and starch in relation to amylose content", *Carbohydr. Polym.*, **2008**, *71*, 380-387.
29. H. Jacobs and J. A. Delcour, "Hydrothermal modifications of granular starch, with retention of the granular structure: A review", *J. Agric. Food Chem.*, **1998**, *46*, 2895-2905.
30. G. Williamson, N. J. Belshaw, D. J. Self, T. R. Noel, S. G. Ring, P. Cairns, V. J. Morris, S. A. Clark and M. L. Parker, "Hydrolysis of A- and B-type crystalline polymorphs of starch by α-amylase, β-amylase and glucoamylase 1", *Carbohydr. Polym.,* **1992**, *18*, 179-187.

Maejo Int. J. Sci. Technol. **2012**, *6*(02), 272-281

Maejo International Journal of Science and Technology

ISSN 1905-7873
Available online at www.mijst.mju.ac.th

Full Paper

Bootstrap confidence intervals for the process capability index under half-logistic distribution

Wararit Panichkitkosolkul [1,*] and Luckhana Saothayanun [2]

[1] Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Phathum Thani, 12121, Thailand
[2] Department of Applied Statistics, Faculty of Science, University of the Thai Chamber of Commerce, Bangkok, 10400, Thailand

* Corresponding author, e-mail: wararit@mathstat.sci.tu.ac.th

Received: 21 August 2011 / Accepted: 26 July 2012 / Published: 27 July 2012

Abstract: This study concerns the construction of bootstrap confidence intervals for the process capability index in the case of half-logistic distribution. The bootstrap confidence intervals applied consist of standard bootstrap confidence interval, percentile bootstrap confidence interval and bias-corrected percentile bootstrap confidence interval. Using Monte Carlo simulations, the estimated coverage probabilities and average widths of bootstrap confidence intervals are compared, with results showing that the estimated coverage probabilities of the standard bootstrap confidence interval get closer to the nominal confidence level than those of the other bootstrap confidence intervals for all situations.

Keywords: bootstrap confidence interval, process capability index, half-logistic distribution.

INTRODUCTION

Balakrishnan [1] introduced the half-logistic distribution as the distribution of the absolute logistic random variable－that is, if Y is a logistic random variable, then $X=|Y|$ has a half-logistic distribution. The probability density function ($f(x)$) and the cumulative distribution function ($F(x)$) are given by

$$f(x)=\frac{2\exp\{-(x-\mu)/\sigma\}}{\sigma\left[1+\exp\{-(x-\mu)/\sigma\}\right]^{2}}, \tag{1}$$

and

$$F(x)=\frac{1-\exp\{-(x-\mu)/\sigma\}}{\left[1+\exp\{-(x-\mu)/\sigma\}\right]}, \quad x>\mu, \quad \sigma>0, \tag{2}$$

where μ and σ are the location and the scale parameters respectively. The graph of the probability density function for half-logistic distribution is shown in Figure 1. The mean and the variance of X are defined as

$$E(X)=\mu+\sigma\ln(4) \quad \text{and} \quad Var(X)=\sigma^2\left[\frac{\pi^2}{3}-(\ln(4))^2\right].$$

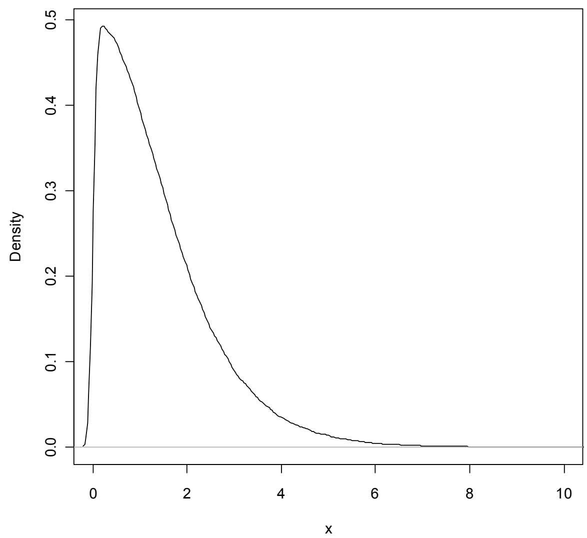


Figure 1. Probability density function for half-logistic distribution with $\mu=0$ and $\sigma=1$

The half-logistic distribution has been widely used for many applications. For example, Balakrishnan [1] has suggested the usage of this distribution as a possible lifetime model with an increasing hazard rate. In addition, Balakrishnan and Chan [2] have shown that the failure times of air conditioning equipment in a Boeing 720 airplane fit the half-logistic distribution quite well. This distribution was also applied to environmental and sports records data [3]. In recent papers, several authors have applied the half-logistic distribution under progressive Type-II censoring [4-6]. As mentioned above, it is known that the half-logistic distribution is an increasing failure rates model with considerable importance in quality control and reliability studies [7-9].

In product quality control, the process capability index (PCI) has been widely adopted as a useful tool. Several process capability indices have been proposed to numerically measure whether a process is capable of manufacturing products that meet customer requirements or specifications [10]. Even though there are many process capability indices, the two most commonly used indices are C_p and C_{pk} [11-12]. The more popular process capability index C_{pk} can be defined as follows [11]:

$$C_{pk}=\min\left\{\frac{USL-\mu}{3\sigma},\frac{\mu-LSL}{3\sigma}\right\}, \tag{3}$$

where USL and LSL denote the upper and lower specification limits of the process respectively, σ is the process standard deviation, and μ is the process mean. As the process standard deviation and the process mean are unknown, they must be estimated from the sample data $\{X_1,...,X_n\}$. The sample mean $\bar{X}$; $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ and the sample standard deviation S; $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ are used to estimate the unknown parameters μ and σ respectively in Eq.(3). The estimator of the process capability index C_{pk} therefore is

$$\tilde{C}_{pk} = \min\left\{\frac{USL - \bar{X}}{3S}, \frac{\bar{X} - LSL}{3S}\right\}.$$

However, the underlying process distribution is non-normal in some situations. Hence it may be a skewed distribution. To deal with these situations, Clements [13] proposed a new method for computing the estimator of the process capability index C_{pk} when the process distribution is non-normal. This estimator is defined as

$$\breve{C}_{pk} = \min\left\{\frac{USL - M}{U_p - M}, \frac{M - LSL}{M - L_p}\right\}, \tag{4}$$

where U_p, L_p and M denote the 99.865$^{\text{th}}$, 0.135$^{\text{th}}$ and 50$^{\text{th}}$ percentiles of the distribution respectively. The advantage of $\breve{C}_{pk}$ shown in Eq.(4) is that it can be applied to any distribution. Kantam et al. [8] discussed the relationship between $\breve{C}_{pk}$ and the probability of a product falling outside the specification limits. When X has a half-logistic distribution, this probability is given by

$$P_t = P(X \le LSL) + P(X \ge USL) = 1 + F(LSL) - F(USL),$$

where $F(\cdot)$ is the cumulative distribution function of a half-logistic distribution shown in Eq.(2). In the case of standard half-logistic distribution, i.e. $\mu = 0$, $\sigma = 1$ in Eq.(1), the values of U_p, L_p and M are 7.300122639, 0.002700002 and $\ln(3) \approx 1.098612289$ respectively. On the other hand, if a scale parameter σ is introduced and known, i.e. $\mu = 0$, $\sigma \ne 1$ in Eq.(1), the optimal estimator of C_{pk} is given by [8]:

$$\breve{C}'_{pk} = \min\left\{\frac{USL - \sigma M}{\sigma(U_p - M)}, \frac{\sigma M - LSL}{\sigma(M - L_p)}\right\}.$$

In practice, the scale parameter σ is unknown. Therefore, the unknown σ must be estimated by its estimator. In this paper the method of moments is used for calculating this estimator. The estimator of C_{pk} for a half-logistic distribution therefore is

$$\hat{C}_{pk} = \min\left\{\frac{USL - \hat{\sigma} M}{\hat{\sigma}(U_p - M)}, \frac{\hat{\sigma} M - LSL}{\hat{\sigma}(M - L_p)}\right\}. \tag{5}$$

where $\hat{\sigma}$ is the method of moments estimator of σ given by $\hat{\sigma} = \bar{X}/\ln(4)$. The properties of the estimator of C_{pk} with mean squared error (MSE) and absolute of bias (|Bias|) are considered. Using Monte Carlo simulations, the MSE and |Bias| are plotted in Figure 2. If σ is fixed and $n \to \infty$,

$MSE(\hat{C}_{pk}) \to 0$ and $|Bias(\hat{C}_{pk})| \to 0$. Therefore, the estimator of C_{pk} given in Eq.(5) is an approximate estimator in terms of MSE and |Bias|.

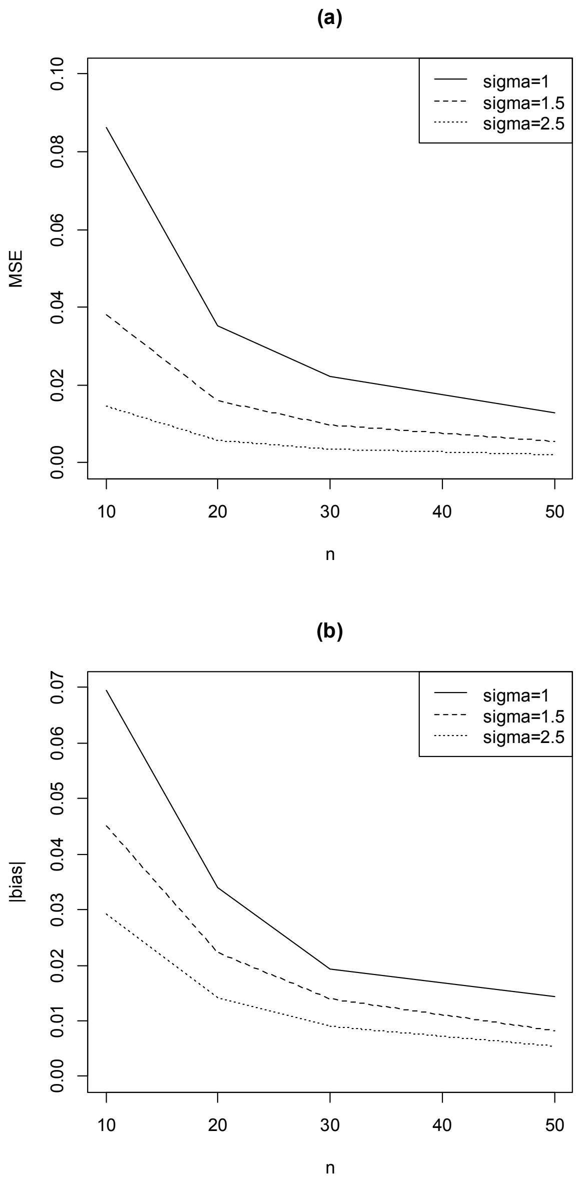


Figure 2. (a) MSE of the estimator of C_{pk} and (b) |Bias| of the estimator of C_{pk} when $\mu = 0$ and $\sigma = 1, 1.5, 2.5$

BOOTSTRAP CONFIDENCE INTERVALS

The bootstrap is a computer-based and resampling method for assigning measures of accuracy to statistical estimates [14]. The advantage of the bootstrap method is that it is a simple approach for estimating biases, standard errors, confidence intervals and so forth for complicated estimators. Furthermore, the distribution of the variable of interest is not mathematically estimated, but rather empirically developed on the characteristics of the distribution of original data [15]. There are many types of bootstrap methods for constructing confidence intervals that have been introduced, e.g. the standard bootstrap method (SB), the percentile bootstrap method (PB) and the bias-corrected percentile bootstrap method (BCPB) [14].

For a sequence of independent and identically distributed random variables, the bootstrap procedure can be defined as follows [14]. Let the random variables $\{X^*_{n,j}, 1 \le j \le m\}$ be the results from sampling m times with replacement from n observations $X_1,\ldots,X_n$. The random variables $\{X^*_{n,j}, 1 \le j \le m\}$ are called the bootstrap samples from the original data $X_1,\ldots,X_n$. The construction of confidence intervals of the process capability index C_{pk} using bootstrap techniques are described in what follows.

Standard Bootstrap (SB) Confidence Interval

Let X^*_b, where $1 \le b \le B$, be the b^{th} bootstrap sample and let $X^*_1,\ldots,X^*_B$ be the B bootstrap samples. The b^{th} bootstrap estimator of C_{pk} is computed by [14]:

$$\hat{C}^{*(b)}_{pk} = \min\left\{ \frac{USL - \hat{\sigma}^{*(b)}M}{\hat{\sigma}^{*(b)}(U_p - M)}, \frac{\hat{\sigma}^{*(b)}M - LSL}{\hat{\sigma}^{*(b)}(M - L_p)} \right\},$$

where

$$\hat{\sigma}^{*(b)} = \bar{X}^{*(b)} / \ln(4), \quad \text{and} \quad \bar{X}^{*(b)} = m^{-1}\sum_{j=1}^{m} X^*_{n,j}.$$

Thus, the standard bootstrap $(1-\alpha)100\%$ confidence interval is

$$CI_{SB} = \left(\bar{\hat{C}}^*_{pk} - Z_{1-\frac{\alpha}{2}} S^*_c, \bar{\hat{C}}^*_{pk} + Z_{1-\frac{\alpha}{2}} S^*_c \right), \tag{6}$$

where $Z_{1-\alpha/2}$ is a $(1-\alpha/2)^{th}$ quantile of the standard normal distribution $\bar{\hat{C}}^*_{pk} = B^{-1}\sum_{i=1}^{B}\hat{C}^{*(i)}_{pk}$ and

$$S^*_c = \sqrt{\frac{1}{B-1}\sum_{i=1}^{B}\left(\hat{C}^{*(i)}_{pk} - \bar{\hat{C}}^*_{pk}\right)^2}.$$

Percentile Bootstrap (PB) Confidence Interval

The percentile bootstrap $(1-\alpha)100\%$ confidence interval is given by [14]:

$$CI_{PB} = \left(\hat{C}^*_{pk\left(\frac{\alpha}{2}B\right)}, \hat{C}^*_{pk\left(\left(1-\frac{\alpha}{2}\right)B\right)} \right), \tag{7}$$

where $\hat{C}^*_{pk(r)}$ is the r^{th} ordered value on the list of the B bootstrap estimator of C_{pk}.

Bias-Corrected Percentile Bootstrap (BCPB) Confidence Interval

The bootstrap distributions obtained using only a sample of the complete bootstrap distribution may be shifted higher or lower than would be expected. Therefore, this approach has been introduced in order to correct for the potential bias. Firstly, using the ordered distribution of $\hat{C}^*_{pk}$, compute the probability $P_0 = P(\hat{C}^*_{pk} \le \hat{C}_{pk})$. Then, $Z_0 = \Phi^{-1}(P_0)$. Therefore, the percentile of the ordered distribution $G^*(\hat{C}^*_{pk})$, $P_L = \Phi\left(2Z_0 - Z_{1-\alpha/2}\right)$ and $P_U = \Phi\left(2Z_0 + Z_{1-\alpha/2}\right)$ are obtained, where $\Phi(\cdot)$ is the standard normal cumulative function. Finally, the bias-corrected percentile bootstrap $(1-\alpha)100\%$ confidence interval is defined as follows [14]:

$$CI_{BCPB} = \left(\hat{C}^*_{pk(P_L B)}, \hat{C}^*_{pk(P_U B)}\right), \qquad (8)$$

where $\hat{C}^*_{pk(r)}$ is the r^{th} ordered value on the list of the B bootstrap estimator of C_{pk}.

To study the different confidence intervals, their estimated coverage probabilities and average widths are considered. For each of the methods considered, a $(1-\alpha)100\%$ confidence interval denoted by (L,U) is obtained (based on $M = 10,000$ replicates). The estimated coverage probability and the average width are given by [16]:

$$\text{Coverage Probability} = \frac{\#(L \le C_{pk} \le U)}{M},$$

and

$$\text{Average Width} = \frac{\sum_{i=1}^{M}(U_i - L_i)}{M}.$$

In the following section, the simulation results are presented in order to evaluate the performance of the confidence intervals CI_{SB}, CI_{PB}, and CI_{BCPB} based on their estimated coverage probabilities and average widths.

SIMULATION RESULTS

In the following, the bootstrap confidence intervals CI_{SB}, CI_{PB} and CI_{BCPB} are compared via Monte Carlo simulation. Using R statistical software [17-19], the data sets are generated from a half-logistic distribution given by Eq.(1) with μ = 0, σ = 1.0, 1.5 and 2.5. The scope of the simulation is set under the sample sizes n = 10, 20, 30, 50 and 100, and LSL and USL are taken as $LSL = 1$ and $USL = 29$. To obtain the estimated coverage probabilities and average widths, the 90% and 95% confidence levels are computed by drawing 1,000 bootstrap samples of sizes $m = n$. The simulation results are summarised in Tables 1-2. As expected, the results show that the estimated coverage probabilities for all confidence intervals get closer to the nominal confidence level with increasing sample sizes n. Likewise, the average widths of all confidence intervals get shorter when n increases. This is intuitive in nature because as n increases, it is possible to estimate the scale parameter σ more accurately. A more interesting result is that the estimated coverage probabilities of the CI_{SB} get closer to the nominal confidence level than those of CI_{PB} and CI_{BCPB}. For example, the estimated coverage probabilities attained by the CI_{SB}, CI_{PB}, and CI_{BCPB} are 0.9451, 0.9335 and 0.9341 respectively, for $n = 50$ and $\sigma = 1.0$. Consequently, the average widths of CI_{SB} are longer than those of CI_{PB} and CI_{BCPB} for all situations.

Table 1. Estimated coverage probabilities and average widths of 90% bootstrap confidence intervals of the process capability index

n	σ	Coverage probability			Average width		
		SB	PB	BCPB	SB	PB	BCPB
10	1.0	0.8926	0.8250	0.8310	0.9829	0.9098	0.8730
	1.5	0.8892	0.8206	0.8222	0.6524	0.6052	0.5813
	2.5	0.9102	0.8335	0.8057	0.3937	0.3638	0.3421
20	1.0	0.8992	0.8634	0.8650	0.6150	0.5962	0.5819
	1.5	0.8942	0.8630	0.8645	0.4095	0.3969	0.3873
	2.5	0.8906	0.8555	0.8524	0.2471	0.2395	0.2331
30	1.0	0.8958	0.8713	0.8734	0.4853	0.4754	0.4674
	1.5	0.8890	0.8669	0.8674	0.3248	0.3182	0.3128
	2.5	0.8934	0.8695	0.8705	0.1947	0.1908	0.1876
50	1.0	0.9015	0.8863	0.8861	0.3675	0.3629	0.3592
	1.5	0.8932	0.8781	0.8789	0.2463	0.2432	0.2407
	2.5	0.8932	0.8730	0.8778	0.1467	0.1449	0.1434
100	1.0	0.9003	0.8930	0.8922	0.2565	0.2546	0.2532
	1.5	0.8983	0.8905	0.8900	0.1709	0.1696	0.1687
	2.5	0.9007	0.8932	0.8928	0.1024	0.1017	0.1012

Table 2. Estimated coverage probabilities and average widths of 95% bootstrap confidence intervals of the process capability index

n	σ	Coverage probability			Average width		
		SB	PB	BCPB	SB	PB	BCPB
10	1.0	0.9357	0.8757	0.8809	1.1651	1.1234	1.0788
	1.5	0.9428	0.8806	0.8818	0.7817	0.7526	0.7234
	2.5	0.9508	0.8881	0.8602	0.4689	0.4495	0.4222
20	1.0	0.9438	0.9102	0.9121	0.7325	0.7219	0.7049
	1.5	0.9386	0.9078	0.9091	0.4878	0.4806	0.4693
	2.5	0.9442	0.9170	0.9143	0.2919	0.2874	0.2799
30	1.0	0.9448	0.9247	0.9256	0.5829	0.5768	0.5672
	1.5	0.9458	0.9215	0.9238	0.3859	0.3819	0.3756
	2.5	0.9424	0.9220	0.9212	0.2321	0.2296	0.2257
50	1.0	0.9451	0.9335	0.9341	0.4396	0.4364	0.4318
	1.5	0.9493	0.9358	0.9373	0.2930	0.2907	0.2876
	2.5	0.9475	0.9327	0.9331	0.1758	0.1744	0.1726
100	1.0	0.9470	0.9393	0.9377	0.3055	0.3038	0.3021
	1.5	0.9509	0.9409	0.9428	0.2037	0.2025	0.2014
	2.5	0.9509	0.9419	0.9428	0.1222	0.1215	0.1209

ILLUSTRATIVE EXAMPLE

To illustrate the bootstrap confidence intervals of the process capability index developed in Section 2, a simulated example is presented. The random samples of sizes $n = 20$ are generated from the half-logistic distribution with $\mu = 0$ and $\sigma = 1$. In this case, we set $LSL = 1$, $USL = 29$, and the true process capability index, C_{pk} = -1/3. The random samples generated are:

0.04	0.14	0.19	0.20	0.23	0.44	0.75	0.81	0.88	1.07
1.07	1.09	1.29	1.50	1.62	1.83	1.91	3.56	5.04	5.15.

In addition, the density plot of the generated samples is displayed in Figure 3. Assuming the half-logistic distribution for the corresponding random samples, three bootstrap confidence intervals of the process capability index with a confidence level of 95% are constructed, and they are shown in Table 3. The value of the true C_{pk} lies in the bootstrap confidence intervals. Additionally, the widths of the confidence intervals are similar to the simulation results.

Table 3. Bootstrap confidence intervals and their widths for the process capability index

Method	Confidence interval	Width
SB	(-0.3608 , 0.5187)	0.8795
PB	(-0.4499 , 0.4111)	0.8610
BCPB	(-0.4448 , 0.4129)	0.8577

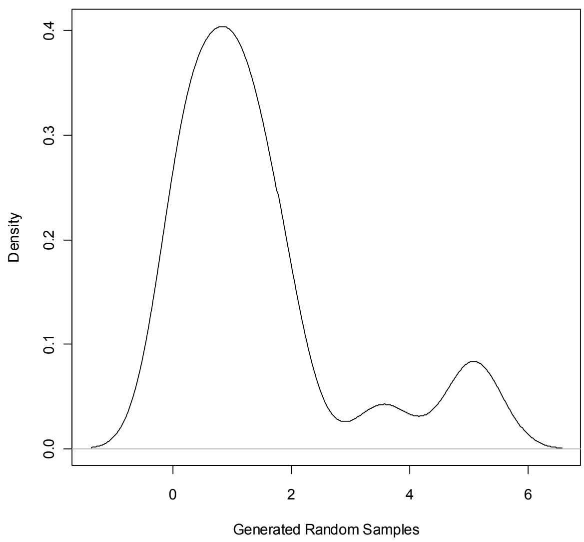


Figure 3. Density plot of generated random samples

CONCLUSIONS

The bootstrap confidence intervals of the process capability index for half-logistic distribution have been proposed. The following were considered: the standard bootstrap confidence interval, the percentile bootstrap confidence interval and the bias-corrected percentile bootstrap confidence interval. By means of Monte Carlo experiments, the performance of the bootstrap prediction intervals was compared by considering their coverage probabilities and average widths. Based on simulation study, the standard bootstrap confidence interval achieved better coverage probability than the other confidence intervals. Thus, the standard bootstrap confidence interval is more appropriate than its counterparts in this setting.

ACKNOWLEDGEMENTS

The authors are very appreciative of the valuable insights and comments provided by the three anonymous referees, which led to many improvements to the paper. The authors also acknowledge the excellent suggestions given by Professor Dr. R.R.L. Kantam, Professor Dr. David E. Giles and Associate Professor Dr. Punnzeyar Patthanangkoor.

REFERENCES

1. N. Balakrishnan, "Order statistics from the half logistic distribution", *J. Stat. Comput. Simul.*, **1985**, *20*, 287-309.
2. N. Balakrishnan and P. S. Chan, "Estimation for the scale half logistic distribution under Type II censoring", *Comput. Stat. Data Anal.*, **1992**, *13*, 123-141.
3. A. K. Mbah and C. P. Tsokos, "Record values from half logistics and inverse Weibull probability distribution functions", *Neural Parallel Sci. Comput.*, **2008**, *16*, 73-92.
4. S. B. Kang, Y. S. Cho and J. T. Han, "Estimation for the half logistic distribution under progressive Type-II censoring", *Commun. Korean Stat. Soc.*, **2008**, *15*, 815-823.
5. N. Balakrishnan and H. M. Saleh, "Relations for moments of progressively Type-II censored order statistics from half-logistic distribution with applications toiInference", *Comput. Stat. Data Anal.*, **2011**, *55*, 2775-2792.
6. D. H. Jang, J. Park and C. Kim, "Estimation of the scale parameter of the half-logistic distribution with multiply type II censored sample", *J. Korean Stat. Soc.*, **2011**, *40*, 291-301.
7. R. R. L. Kantam and K. Rosaiah, "Half logistic distribution in acceptance sampling based on life tests", *IAPQR Trans.*, **1998**, *23*, 117-125.
8. R. R. L. Kantam, K. Rosaiah and R. Subba Rao, "Estimation of process capability index for half logistic distribution", *Int. Trans. Math. Sci. Comput.*, **2010**, *3*, 61-66.
9. R. R. L. Kantam, K. Rosaiah and M. S. R. Anjaneyula, "Estimation of reliability in multicomponent stress-strength model: Half logistic distribution", *IAPQR Trans.*, **2000**, *25*, 43-52.
10. S. Niwitpong and J. Sappakitkamjorn, "Biased estimation of the capability indices C_p, C_{pk} and C_{pm} using bootstrap and jackknife methods", *Asia Pac. Manag. Rev.*, **2006**, *10*, 1-7.
11. V. E. Kane, "Process capability indices", *J. Qual. Technol.*, **1986**, *18*, 41-52.
12. J. Zhang, "Conditional confidence intervals of process capability indices following rejection of preliminary tests", Ph.D. Thesis, **2010**, The University of Texas at Arlington, USA.
13. J. A. Clements, "Process capability calculations for non-normal distributions", *Qual. Prog.*, **1989**, *22*, 95-100.

14. B. Efron and R. J. Tibshirani, "An Introduction to the Bootstrap", Chapman and Hall, New York, **1993**.
15. N. Péladeau and Y. Lacouture, "SIMSTAT: Bootstrap computer simulation and statistical program for IBM personal computers", *Behav. Res. Methods*, **1993**, *25*, 410-413.
16. W. Panichkitkosolkul, "A simulation comparison of new confidence intervals for the coefficient of variation of a Poisson distribution", *Silpakorn U. Sci. Technol. J.*, **2010**, *4*, 14-20.
17. R. Ihaka and R. Gentleman, "R: A language for data analysis and graphics", *J. Comput. Graph. Statist.*, **1996**, *5*, 299-314.
18. W. N. Venables, D. M. Smith and the R Core Team, "An Introduction to R: A Programming Environment for Data Analysis and Graphics, Version 2.15.1", R Core Team, **1999-2012** (Online at: http://cran.r-project.org/doc/manuals/R-intro.pdf).
19. The R Core Team, "R: A Language and Environment for Statistical Computing", R Foundation for Statistical Computing, **1999-2010** (Online at: http://cran.r-project.org/doc/manuals/refman.pdf).

Maejo Int. J. Sci. Technol. **2012**, *6*(02), 282-296

Maejo International Journal of Science and Technology

ISSN 1905-7873
Available online at www.mijst.mju.ac.th

Full Paper

A multithreaded scheduling model for solving the Tower of Hanoi game in a multicore environment

Alaa M. Al-Obaidi * and Sai Peck Lee

Department of Software Engineering, Faculty of Computer Science & Information Technology, University of Malaya, Kuala Lumpur, Malaysia

* Corresponding author, e-mail: alaa@siswa.um.edu.my

Received: 10 May 2011 / Accepted: 1 August 2012 / Published: 7 August 2012

Abstract: Modern computer systems greatly depend on multithreaded scheduling to balance the workload among their working units. One of the multithreaded scheduling techniques, the work-stealing technique has proven effective in balancing the distribution of threads by stealing threads from the working cores and reallocating them to the non-working cores. In this study, we propose a new strategy that extends the work-stealing technique by enabling it to select the richest core prior to any redistribution process. In order to obtain practical results, we applied this new strategy of balancing threads to one of the divide-and-conquer problems, the Tower of Hanoi game. A multithreaded scheduling model which is a hierarchical model was designed to work under the control of this new strategy. A modelling tool was used to simulate and verify the designed model. The proposed model was shown by the simulation process to exhibit consistency and stability in reaching the desired result. Scalability, concurrency, simplicity and fair load distribution among the modelled cores are the main beneficial characteristics of this model.

Keywords: multithreaded scheduling, Tower of Hanoi, work-stealing technique, divide-and-conquer problem, coloured Petri nets

---

INTRODUCTION

The general trend in the multicore industry is to increase the number of cores per chip. The rapid growth in the number of cores per chip imposes new requirements for software designers who need to make their products more adaptable with the new developments taking place on the hardware side of the industry. In order to address this challenge, multithreaded scheduling techniques have been developed to provide a solution for managing the continuing increase in the number of cores. These techniques are in charge of assigning and distributing the workload among the cores. For instance, the work-stealing scheduling technique has proven effective in balancing the distribution of threads in multicore environments. The technique aims to balance the distribution of threads by stealing threads from the working cores and reallocating them to the non-working cores [1-9].

However, in view of the increasing number of cores, the work-stealing technique may not be sufficient to meet the demands of multicore technology. Hence, motivated by the need to meet these demands, the development of new software strategies that are built on the basis of the work-stealing technique is the main aim of this study. We aim to design a multithreaded scheduling model that improves the work-stealing technique by enabling it to select the richest core (i.e. the one that has the most threads) when balancing thread distribution among working and non-working cores.

In this paper, we present a new strategy for balancing the distribution of threads. The richest selection scheduling (RSS) strategy extends the work-stealing technique so that it can choose the core that has the most threads from among the available cores when there is a need to redistribute threads. We have designed a multithreaded scheduling model, which is a hierarchical model that consists of two types of schedulers. The first scheduler is called a High-Level Scheduler, which is in charge of applying the RSS strategy. This scheduler was implemented through designing a new algorithm, the RSS algorithm. The algorithm is in charge of making all the cores work concurrently by stealing a certain number of threads from the working (victim) cores and reallocating them to the non-working (thief) cores. Where two or more victim cores are available, the richest core (the one that has the most threads) is preferred. The second scheduler is called the Core Scheduler. This scheduler is in charge of thread creation and calculation. It is also used to create and calculate the moves for the Tower of Hanoi game. A new algorithm, the Tower of Hanoi multithreaded scheduling (THMS) algorithm was designed for this purpose. We applied these algorithms by using coloured Petri net (CPN) [10-13] as the modelling language and coloured Petri nets meta language (CPN-ML) [14-16] as the coding language. We used CPN-Tool [17] as a software tool to enable us to create, simulate and verify the correctness of the designed model.

The Tower of Hanoi

The general idea behind any divide-and-conquer problem can be summarised as the continuous partitioning of a given problem till a certain condition is achieved [18-19]. After that, the conquering part is executed in a manner that depends on the nature of the problem. The Tower of Hanoi computer game is based on a puzzle that was first published by a French mathematician (François Éduoard Anatole Lucas) in 1883 [18]. The game consists of three pillars and n disks. Initially, two of the pillars are empty. The first pillar contains n disks stacked with the largest disk at the bottom. Figure 1 shows an example of this game. In this example, three disks are located on the first pillar. The smallest disk has the number 1 while the largest disk has the number 3. The objective of this game is to move all the disks one by one from pillar 1 to pillar 3 under one condition: putting a large disk on top of a small one is not allowed. The output of this game is represented by a sequence of moves. Any single move consists of three parameters: disk number, source pillar and destination pillar. The number of steps is equal to $(2^n) - 1$.

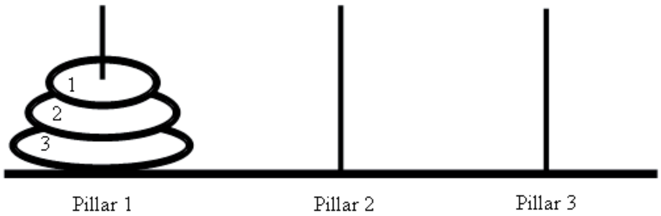


Figure 1. The Tower of Hanoi pillars with three disks

Prior to the development of multicore technology, the solution to such problem had to be done serially. That is, the generation of the steps had to be made one by one. However, with the advent of multicore technology, solutions can be done faster through multithreading and concurrency techniques [20]. However, the modelling of multithreaded concurrent systems represents a great challenge due to the non-deterministic nature of such systems, in addition to the difficulty of thread synchronisation The divide-and-conquer technique, concurrency, and multithreaded scheduling have the following in common: the main problem can be divided into several parts, and each part can be assigned to a thread. The ability to allocate a core for each thread makes all the threads work concurrently.

RELATED WORK

The main idea of work stealing is attributed to Blumofe and Leiserson [2]. They designed an algorithm that is able to schedule well-structured multithreaded computations. Although the result was good, the algorithm could not deal with the new environments where multiprogramming is used. This is due to the design of the algorithm's mechanism which deals with a fixed set of processors. Arora et al. [3] made an improvement by designing an algorithm that can deal with a multiprogrammed environment instead of a dedicated one. However, their algorithm encountered some problems with respect to memory management. Overflow easily occurred due to the use of arrays in representing deques. The sizes of the arrays had to be adjusted many times [4]. Several improvements have been built on the development made by Arora et al., e.g. 'stealing the half', a new idea that was introduced by Hendler et al. [5]. As the name implies, half of the threads can be stolen in one trial. Two other contributions based on Arora et al. are a locality-guided work-stealing algorithm that improves the data locality of multithreaded computations by allowing a thread to have an affinity for a processor suggested by Acar et al. [6], and a simple lock-free work-stealing technique proposed by Chase and Lev [7]. Chase and Lev's algorithm is based on using a cyclic array that can easily deal with overflows; memory size is the only limitation to their algorithm.

We have developed a scheduling strategy which has been designed on the principle of work stealing and is simple and direct [8-9]. In this strategy, which was applied to solve two of the divide-and-conquer problems, namely Fibonacci Series and Binary Search, the search for the victim cores is simply achieved through investigating each core starting from core number 1 to the end of the core list. Although this technique succeeds in balancing thread distribution in a direct and simple way, it lacks selection efficiency. In certain cases, the scheduler may pick a victim core that has too few threads compared with other victim cores. This leads to unnecessary repetitions of the distribution process because the process itself may not satisfy the needs of the thief cores.

CPN and CPN-Tool

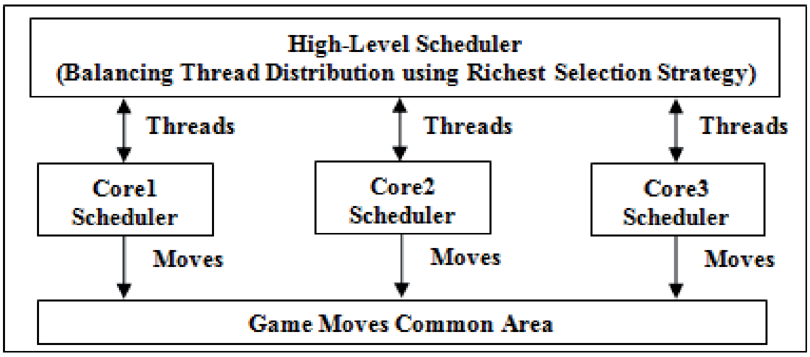
CPN is a graphical discrete-event language designed to model and validate concurrent systems [10-13]. CPN has been developed from Petri nets [21-22], the main difference between Petri nets and CPN being the addition of types to CPN as well as the ability to write expressions and functions in standard meta language (SML) [14-16]. The CPN model is an executable model in the sense that the process of execution shows the different states of the system represented by the model.

CPN-Tool is a software tool that is designed to create, simulate and validate CPN [17]. CPN-Tool is a GUI tool that provides all the interaction methods such as menus and toolbars in addition to giving feedback messages when errors are encountered during the process of checking the syntax of codes. The ability to execute models is one of the main advantages of this tool. In addition, the tool supports hierarchical modelling, which simplifies complicated designs. Additionally, CPN-Tool provides a monitoring mechanism which permits observation of the behaviour of the elements of the model. CPN-Tool uses CPN-ML for writing declarations, expressions and codes [17]. CPN-ML is a language that can be used to write net inscriptions. These inscriptions include expressions on the arcs, codes that control transitions as well as declarations of the types and variables that are included in the CPN model. The CPN-ML has been built based on SML [14-16]. The tool has proved successful in the world of modelling [17].

DESIGN METHODOLOGY

In this study, we propose a concurrent multithreaded scheduling model that is able to balance thread distribution among a set of modelled cores through applying a new strategy, namely the RSS strategy. In addition, the model is able to schedule the computations of moves in the Tower of Hanoi game. The methodology is based on building two types of schedulers: the High-Level Scheduler and the Core Scheduler (Scheme 1). Basically, there is only one high-level scheduler while there are as many core schedulers as the number of modelled cores (three in the case of Scheme 1). The High-Level Scheduler exchanges threads with the Core Schedulers while the latter are in charge of creating these threads in addition to generating game moves and keeping them in a common area.

The thread is designed as a 7-tuple: *ThreadId*, *FatherId*, *Disk-No*, *Order*, *Source*, *Through* and *Destination*, the first two representing the thread's identifier and the thread's father identifier respectively. *ThreadId* and *FatherId* are denoted as x (x = a positive integer). The third parameter, *Disk-No*, holds the number of the disk. The fourth parameter, *Order*, represents a sequence number that is reserved for the game moves. The last three parameters, *Source*, *Through* and *Destination*, represent the numbers of the pillars.



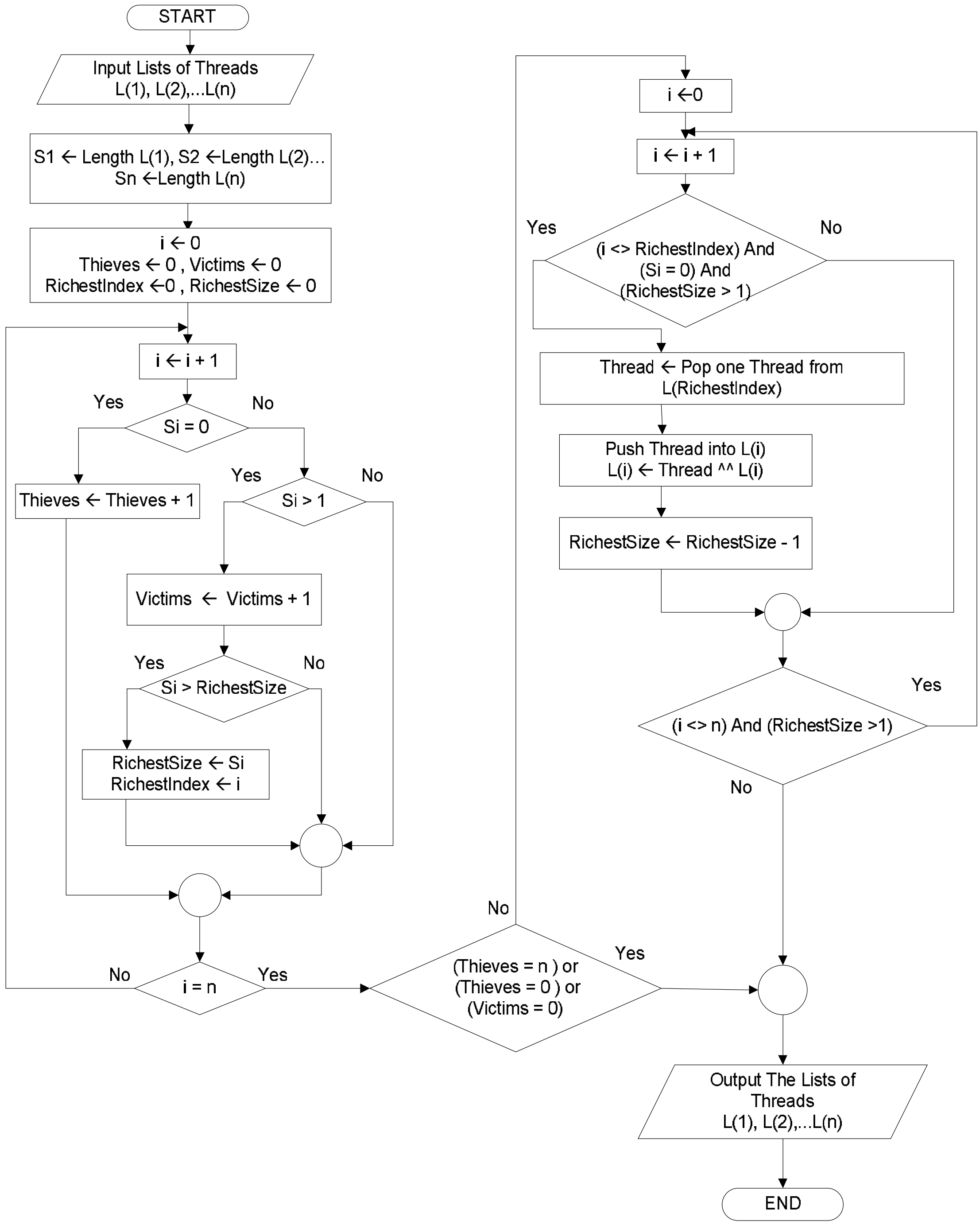
Scheme 1. Scheme of the model with three cores

The game move is modelled as a 4-tuple: *Order*, *Disk-No*, *Source* and *Destination*. It holds the information of moving a single disk numbered as *Disk-No* from pillar number *Source* to pillar number *Destination*. The *Order* parameter is necessary for putting all the moves in order in the common area. Since each modelled core concurrently generates its own set of moves, it becomes necessary to add a parameter that controls the arranging of these moves.

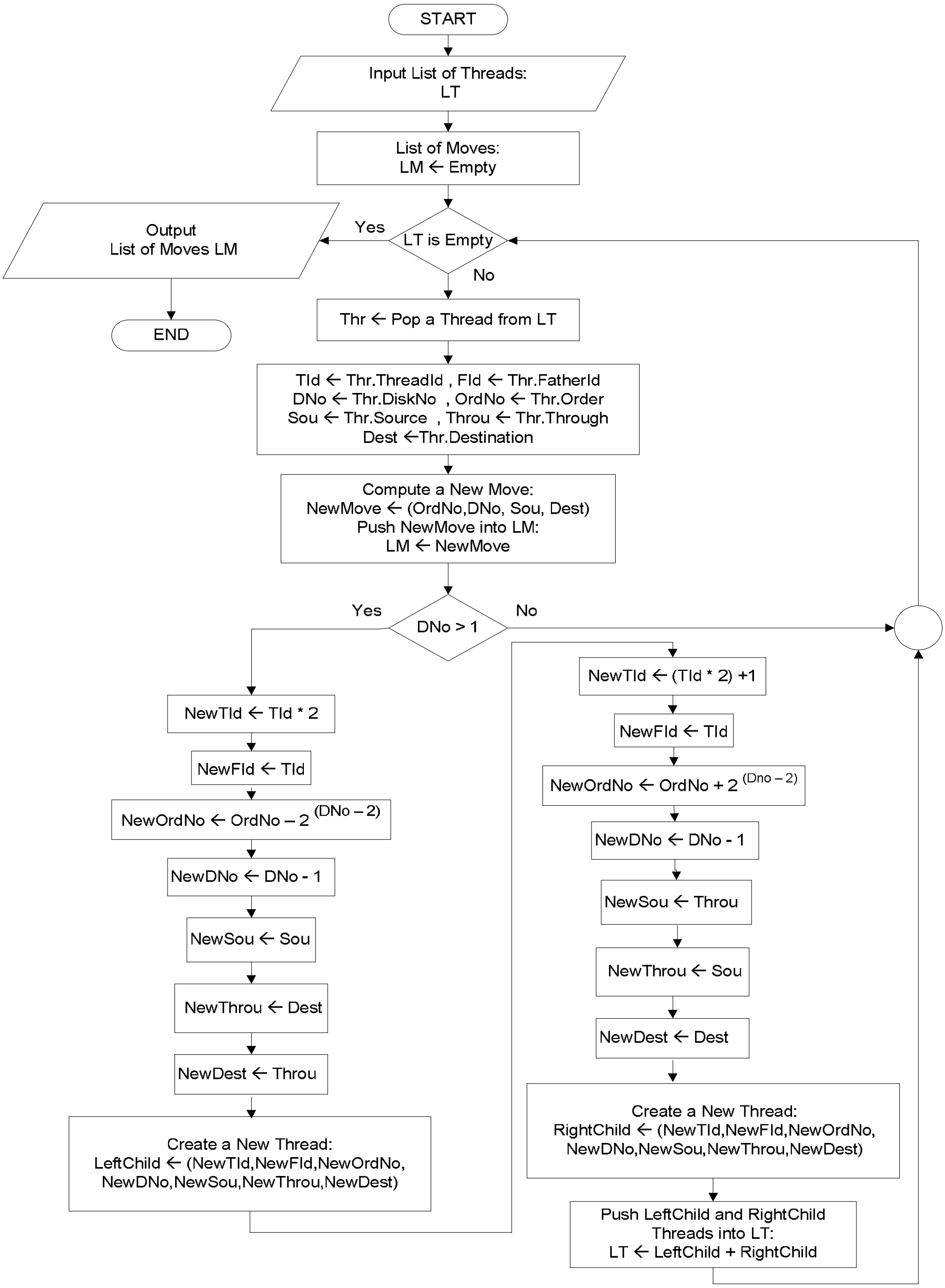
The High-Level Scheduler (Scheme 2) has the role of controlling thread distribution among the modelled cores. The process of distribution can be achieved through stealing threads from the working (victim) core that has the highest number of threads and then reallocating those threads to a set of non-working (thief) cores.

The mechanism of the Core Scheduler, as shown in Scheme 3, creates a binary tree of threads and game moves. The scheduler divides any given thread into left and right descendant threads and the division process continues until no thread can be divided. The number 0 is chosen as the *Order* number of the main thread. Therefore, the *Order* numbers of the left-thread children will be less than their counterparts on the right side. Thus, some threads' identifiers have negative *Order* numbers (in CPN-ML, the negative sign is ~) and the other threads' identifiers have positive *Order* numbers (an example being given in Figure 5). The scheduler also generates game moves; a game move is configured from the thread itself. It is an abstract representation of the thread by holding the *Disk-No* and the numbers of the *Source* and *Destination* pillars. All the cores' schedulers keep their game moves in a common area, as shown in Scheme 1. The schedulers are independent; they can work concurrently to exploit the cores to reduce the overall execution time.

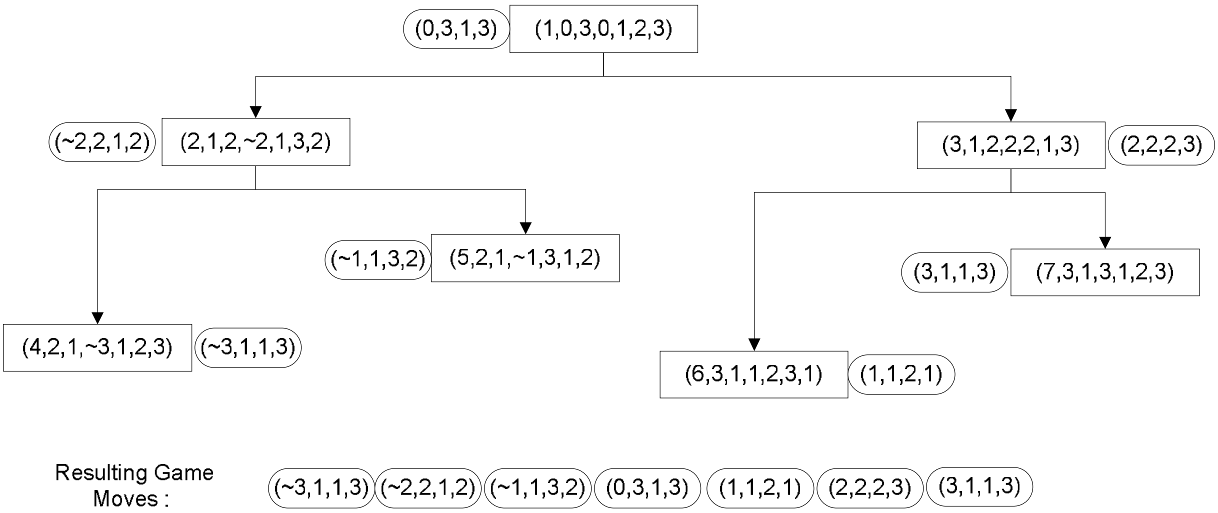
Scheme 4 shows an example of a binary tree of threads and game moves for a model that has three disks. Threads and game moves are included inside rectangles and ellipses respectively. The first resulting game move consists of moving disk 1 from pillar 1 to pillar 3; the second move consists of moving disk 2 from pillar 1 to pillar 2; etc. The first parameter in each move is used to put the generated move in its proper order.



Scheme 2. High-Level Scheduler mechanism – RSS strategy



Scheme 3. Core Scheduler mechanism – scheduling of moves in the Tower of Hanoi game



Scheme 4. Example of scheduling the threads and moves ofTower of Hanoi game with three disks

Building the CPN Model

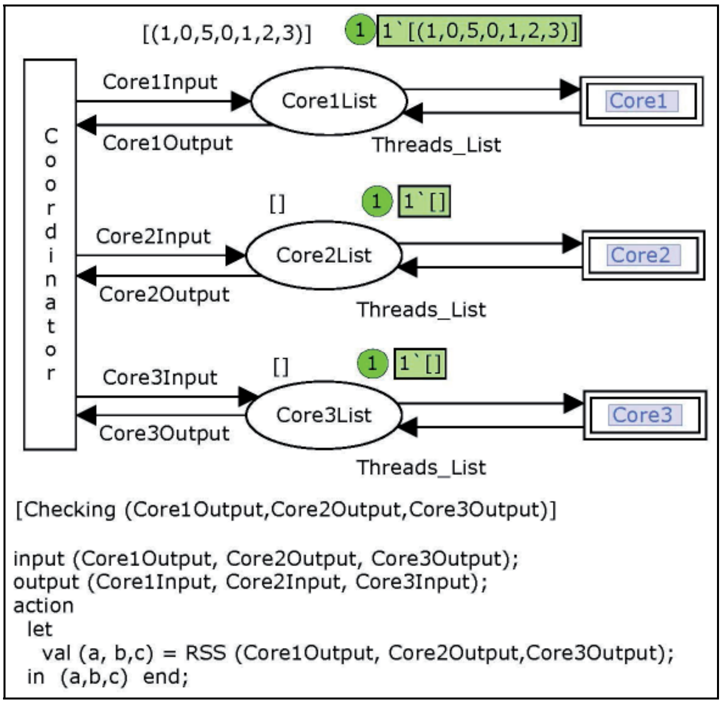
The CPN model of the main page is shown in Scheme 5. The model simulates Scheme 1. It includes three places (Core1List, Core2List and Core3List), one transition (Coordinator) and three substituted transitions (Core1, Core2 and Core3). In CPN a place is an oval shape which holds data. In our model each place holds a list of threads. The place usually is accompanied by the following three items:

1) An initial value which is located on the top left of the place. The place Core1List contains a list with a single thread called the main thread ([(1, 0, 5, 0, 1, 2, 3)]). The first two parameters, i.e. 1 and zero, symbolise *ThreadId* and the thread's *FatherId* respectively. Number 5 represents the number of disks; for example, we are planning to move five disks from pillar 1 to pillar 3. The fourth parameter, zero, stands for the *Order* parameter of the first (main) thread. The remaining three parameters represent the pillar numbers 1, 2 and 3. The cores, Core2List and Core3List, have the initial value '[]', which means that the cores initially have no threads (empty list of threads).

2) A current value which is located on the top right of the place. The current values are frequently changed during the simulation process while the initial values never change. A current value is symbolised by a circle and a rectangle. The rectangle displays the value while the circle shows the number of values. In the case of the list of threads, we have one value, that is, a single list of threads.

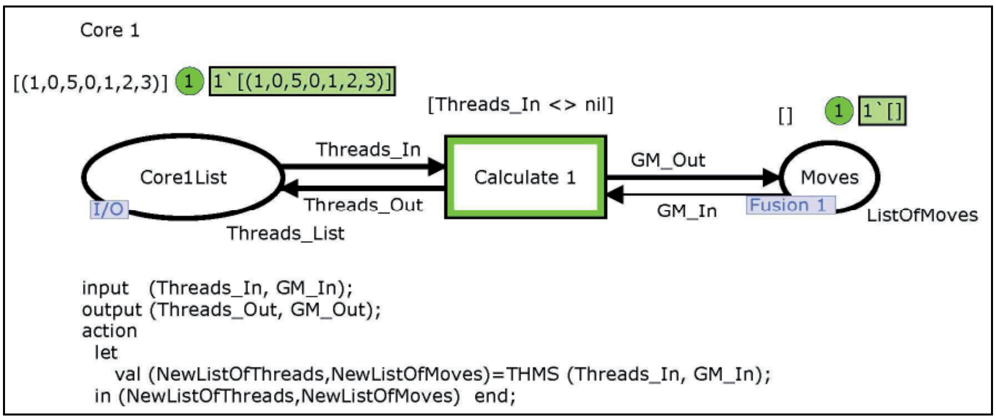
3) A data type *Threads-List* which is located on the bottom right of the place. This data type is designed in CPN-ML to indicate the type of data inside the place.

In CPN, transitions represent the action units. The transition Coordinator is in charge of executing high-level scheduling (described in Scheme 2). To activate the transition Coordinator, the Checking function should return a Boolean true value. This function works as a guard; it returns true only if there is at least one victim and at least one thief. The code below the Checking function represents instructions that will be carried out when transition Coordinator is executed by the CPN-Tool's simulator. The transition reads the lists of the cores through Core1Output, Core2Output and



Scheme 5. CPN model of the main page

Core3Output, updates the lists and then sends the feedback as Core1Input, Core2Input and Core3Input. The transition executes the RSS function to do the updating task. This CPN-ML function has been designed to execute the mechanism of the high-level scheduling strategy, which results in a fair distribution of the threads. Finally, to the right of each place there is a substitute transition: Core1, Core2 and Core3. Each substitute transition corresponds to a core scheduler (Scheme 1). The structure of a substitute transition is illustrated in Scheme 6, which represents Core1. Other substitute transitions have the same structure.



Scheme 6. CPN model of Core1

The model in Scheme 6 shows the content of Core1. The core consists of two places: Core1List and Moves, in addition to one transition: Calculate 1. The Calculate-1 transition reads a list of threads (stored in Core1List) which is represented by Threads_In, updates it and sends it back as Threads_Out. The I/O symbol at the lower-left corner of the place Core1List represents a port tag which is a mechanism offered by the CPN-Tool to connect places from different pages. It allows the Coordinator to add/take threads to/from the places. This means that the place Core1List in Scheme 6 is just a copy of the place Core1List in Scheme 5. This kind of hierarchy simplifies the communication between the pages' places in the model.

The Calculate-1 transition also reads a list of ordered game moves represented by GM_In (stored in the place Moves), updates it and then sends it back as GM_Out. The place Moves has an initial value which consists of an empty list '[]'. In addition, the place Moves has a Fusion-1 tag. Fused place is one of the CPN-Tool mechanisms used in creating shared areas [17]. In our example, the place Moves in each core will share the same game moves list. The Calculate-1 transition has a guard expression ([Threads_In <> nil]) located at the top of the transition. This Boolean expression controls the activation of the transition. The expression returns true (that is, allows the transition to be executed) only if the list of threads is not empty. The remaining cores have exactly the same structure as Core1 except that they start with empty lists of threads. The THMS function applies the mechanism shown in Scheme 3 — that is, this function is the CPN-ML copy of the Core Scheduler. The function receives and updates two parameters, a list of threads and a list of game moves.

Thus, both the RSS and THMS functions control the movements of the threads within the elements of the model. The RSS strategy is implemented by the RSS function while generating the threads, and the movements of the game are implemented by the THMS function.

RESULTS OF SIMULATION

CPN-Tool is a single thread tool and it chooses transitions randomly. At every stage, the tool checks the available active transitions. An active transition must fulfil two conditions: first, its input places are not empty; and second, the transition's guard function (if exists) returns true. Next, the tool picks one of the transitions that is ready in order to execute it. If the tool picks the Coordinator transition, this will cause the reallocation of the threads in the entire model (applying the RSS function). However, when the tool picks one of the Calculate transitions, then new threads and game moves are generated (applying the THMS function).

The following represents an example of a simulation process in the model. The main thread is [(1,0,5,0,1,2,3)] located in Core1. We plan to move five disks from pillar 1 to pillar 3 with the help of pillar 2. Initially, the places Core2list and Core3List are empty. The simulation process starts by executing the only active transition, i.e. Calculate 1 in Core1. Table 1 lists all the selected transitions, cores' contents and game moves during the simulation process.

Table 1. List of transitions selected by CPN-Tool during the simulation process in addition to the current contents of the places. (Transitions Calculate 1, Calculate 2, Calculate 3 and Coordinator have been shortened to C1, C2, C3 and Coo respectively.)

Seq.	Transition	Core-1 thread	Core-2 thread	Core-3 thread	Generated move
1	-	(1,0,5,0,1,2,3)	Nil	Nil	Nil
2	C1	(2,1,4,~8,1,3,2),(3,1,4,8,2,1,3)	Nil	Nil	(0,5,1,3)
3	C1	(4,2,3,~12,1,2,3), (5,2,3,~4,3,1,2),(3,1,4,8,2,1,3)	Nil	Nil	(~8,4,1,2)
4	Coo	(3,1,4,8,2,1,3)	(4,2,3,~12,1,2,3)	(5,2,3,~4,3,1,2)	Nil
5	C2	(3,1,4,8,2,1,3)	(8,4,2,~14,1,3,2), (9,4,2,~10,2,1,3)	(5,2,3,~4,3,1,2)	(~12,3,1,3)
6	C1	(6,3,3,4,2,3,1),(7,3,3,12,1,2,3)	(8,4,2,~14,1,3,2), (9,4,2,~10,2,1,3)	(5,2,3,~4,3,1,2)	(8,4,2,3)
7	C2	(6,3,3,4,2,3,1),(7,3,3,12,1,2,3)	(16,8,1,~15,1,2,3), (17,8,1,~13,3,1,2), (9,4,2,~10,2,1,3)	(5,2,3,~4,3,1,2)	(~14,2,1,2)
8	C2	(6,3,3,4,2,3,1),(7,3,3,12,1,2,3)	(17,8,1,~13,3,1,2), (9,4,2,~10,2,1,3)	(5,2,3,~4,3,1,2)	(~15,1,1,3)
9	C1	(12,6,2,2,2,1,3),(13,6,2,6,3,2,1), (7,3,3,12,1,2,3)	(17,8,1,~13,3,1,2), (9,4,2,~10,2,1,3)	(5,2,3,~4,3,1,2)	(4,3,2,1)
10	C2	(12,6,2,2,2,1,3),(13,6,2,6,3,2,1), (7,3,3,12,1,2,3)	(9,4,2,~10,2,1,3)	(5,2,3,~4,3,1,2)	(~13,1,3,2)
11	C1	(24,12,1,1,2,3,1),(25,12,1,3,1,2,3), (13,6,2,6,3,2,1),(7,3,3,12,1,2,3)	(9,4,2,~10,2,1,3)	(5,2,3,~4,3,1,2)	(2,2,2,3)
12	C2	(24,12,1,1,2,3,1),(25,12,1,3,1,2,3), (13,6,2,6,3,2,1),(7,3,3,12,1,2,3)	(18,9,1,~11,2,3,1), (19,9,1,~9,1,2,3)	(5,2,3,~4,3,1,2)	(~10,2,2,3)
13	C2	(24,12,1,1,2,3,1),(25,12,1,3,1,2,3), (13,6,2,6,3,2,1),(7,3,3,12,1,2,3)	(19,9,1,~9,1,2,3)	(5,2,3,~4,3,1,2)	(~11,1,2,1)
14	C2	(24,12,1,1,2,3,1),(25,12,1,3,1,2,3), (13,6,2,6,3,2,1),(7,3,3,12,1,2,3)	Nil	(5,2,3,~4,3,1,2)	(~9,1,1,3)
15	C1	(25,12,1,3,1,2,3),(13,6,2,6,3,2,1), (7,3,3,12,1,2,3)	Nil	(5,2,3,~4,3,1,2)	(1,1,2,1)
16	C1	(13,6,2,6,3,2,1), (7,3,3,12,1,2,3)	Nil	(5,2,3,~4,3,1,2)	(3,1,1,3)
17	C1	(26,13,1,5,3,1,2),(27,13,1,7,2,3,1), (7,3,3,12,1,2,3)	Nil	(5,2,3,~4,3,1,2)	(6,2,3,1)
18	Coo	(27,13,1,7,2,3,1),(7,3,3,12,1,2,3)	(26,13,1,5,3,1,2)	(5,2,3,~4,3,1,2)	Nil
19	C3	(27,13,1,7,2,3,1),(7,3,3,12,1,2,3)	(26,13,1,5,3,1,2)	(10,5,2,~6,3,2,1), (11,5,2,~2,1,3,2)	(~4,3,3,2)
20	C2	(27,13,1,7,2,3,1),(7,3,3,12,1,2,3)	Nil	(10,5,2,~6,3,2,1), (11,5,2,~2,1,3,2)	(5,1,3,2)
21	Coo	(27,13,1,7,2,3,1),(7,3,3,12,1,2,3)	(10,5,2,~6,3,2,1)	(11,5,2,~2,1,3,2)	Nil
22	C2	(27,13,1,7,2,3,1),(7,3,3,12,1,2,3)	(20,10,1,~7,3,1,2), (21,10,1,~5,2,3,1)	(11,5,2,~2,1,3,2)	(~6,2,3,1)
23	C3	(27,13,1,7,2,3,1),(7,3,3,12,1,2,3)	(20,10,1,~7,3,1,2), (21,10,1,~5,2,3,1)	(22,11,1,~3,1,2,3) , (23,11,1,~1,3,1,2)	(~2,2,1,2)
24	C2	(27,13,1,7,2,3,1),(7,3,3,12,1,2,3)	(21,10,1,~5,2,3,1)	(22,11,1,~3,1,2,3) , (23,11,1,~1,3,1,2)	(~7,1,3,2)
25	C3	(27,13,1,7,2,3,1),(7,3,3,12,1,2,3)	(21,10,1,~5,2,3,1)	(23,11,1,~1,3,1,2)	(~3,1,1,3)
26	C2	(27,13,1,7,2,3,1),(7,3,3,12,1,2,3)	Nil	(23,11,1,~1,3,1,2)	(~5,1,2,1)
27	Coo	(7,3,3,12,1,2,3)	(27,13,1,7,2,3,1)	(23,11,1,~1,3,1,2)	Nil
28	C1	(14,7,2,10,1,3,2),(15,7,2,14,2,1,3)	(27,13,1,7,2,3,1)	(23,11,1,~1,3,1,2)	(12,3,1,3)
29	C3	(14,7,2,10,1,3,2),(15,7,2,14,2,1,3)	(27,13,1,7,2,3,1)	Nil	(~1,1,3,2)

Table 1. (Continued)

Seq.	Transition	Core-1 thread	Core-2 thread	Core-3 thread	Generated move
30	C2	(14,7,2,10,1,3,2),(15,7,2,14,2,1,3)	Nil	Nil	(7,1,2,1)
31	Coo	(15,7,2,14,2,1,3)	(14,7,2,10,1,3,2)	Nil	Nil
32	C1	(30,15,1,13,2,3,1), (31,15,1,15,1,2,3)	(14,7,2,10,1,3,2)	Nil	(14,2,2,3)
33	C2	(30,15,1,13,2,3,1), (31,15,1,15,1,2,3)	(28,14,1,9,1,2,3), (29,14,1,11,3,1,2)	Nil	(10,2,1,2)
34	Coo	(30,15,1,13,2,3,1), (31,15,1,15,1,2,3)	(29,14,1,11,3,1,2)	(28,14,1,9,1,2,3)	Nil
35	C2	(30,15,1,13,2,3,1), (31,15,1,15,1,2,3)	Nil	(28,14,1,9,1,2,3)	(11,1,3,2)
36	C3	(30,15,1,13,2,3,1), (31,15,1,15,1,2,3)	Nil	Nil	(9,1,1,3)
37	Coo	(31,15,1,15,1,2,3)	(30,15,1,13,2,3,1)	Nil	Nil
38	C2	(31,15,1,15,1,2,3)	Nil	Nil	(13,1,2,1)
39	C1	Nil	Nil	Nil	(15,1,1,3)

Now if we remove the indices from the moves, we get the following moves: (1,1,3),(2,1,2),(1,3,2),(3,1,3),(1,2,1),(2,2,3),(1,1,3),(4,1,2),(1,3,2),(2,3,1),(1,2,1),(3,3,2),(1,1,3),(2,1,2),(1,3,2),(5,1,3),(1,2,1),(2,2,3),(1,1,3),(3,2,1),(1,3,2),(2,3,1),(1,2,1),(4,2,3),(1,1,3),(2,1,2),(1,3,2),(3,1,3),(1,2,1),(2,2,3),(1,1,3). In the above list, each move is of the following structure: (Disk No, Source Pillar and Destination Pillar).

DISCUSSION

The model was executed several times. All the trials led to the same result, viz. the generation of the correct sequence of moves. However, different trials have different sequences of transition activations. This is due to the non-deterministic nature of the CPNs. Nevertheless, all the trials generate the same number and sequence of game moves.

The RSS strategy can be applied to any divide-and-conquer problem. The strategy is scalable; it can be easily expanded to deal with any number of cores. Moreover, it is an improvement on the In-Order strategy [8-9] with respect to the searching method. The In-Order strategy picks the first encountered victim regardless of the threads' richness. Although simple and direct, the In-Order strategy becomes inactive when it chooses a victim core that has only a few threads, which may happen frequently. On the other hand, the RSS strategy picks the wealthiest core to ensure that a maximum number of threads can be moved to the thief cores.

In Table 1 the simulator needs 39 steps to reach the final list of moves. Stealing happens in steps 4, 18, 21, 27, 31, 34 and 37. Steps 18 and 27 take advantage of the RSS strategy. The stealing in these steps happens based on a search for the richest core, which is preferred. In the other stealing steps (steps 4, 21, 31, 34 and 37) we can see that there is either only one victim core or the victim cores have the same degree of richness. The stealing and the distribution processes obviously consume some time. However, with the steady increase in the number of cores per chip currently taking place, this kind of sacrifice should be accepted in view of the great advantage that we can gain from ensuring that all the cores are as busy as possible.

Concurrency is one of the main features in this study. The cores work concurrently in most of the steps in Table 1. From step 4 to step 39, many steps have the potential to occur at the same time with other steps. These concurrent steps are: (5 and 6), (8 and 9), (10 and 11), (14 and 15), (19 and 20), (22 and 23), (24 and 25), (28, 29, and 30), (32 and 33), (35 and 36), and (38 and 39). Any increase in the number of cores will definitely increase the number of steps that can occur concurrently. On the other hand, it was observed that Core 3 is less active than Core 1 and Core 2. Despite the fact that in every new run the simulation process starts with Core 1, there will be different sequences of active transitions. In other words, the continuous running of the model will generate different percentages of core activation. This is due to the non-deterministic nature of the CPNs, yet all the runs end with the same result.

The simplicity of the design is another feature of the model. It is so easy to add a new core to the model. New cores can be easily copied and pasted in the model. This is one of the advantages of CPN-Tool. In general, the tool can easily support the expansion of the model by replicating the sub-pages as long as the model is hierarchically well designed.

The language of coding, CPN-ML, was found to play a significant role in supporting the mechanism of the model. This is because firstly CPN-ML is free of side effects. This advantage eliminates the problems of side effects present in other studies that use imperative languages. Using CPN-ML makes the behaviour of the model much more understandable. Second, a huge number of built-in functions are included in this language. This makes it easier to deal with structures such as lists which we use to store the threads and the game moves. In addition, as a functional language derived from SML, CPN-ML uses linked lists when creating lists. Compared with other languages, the use of linked lists has two main advantages:

(1) In linked lists there are no problems such as overflow or managing array indices. Previous studies used fixed arrays when storing threads. This led to the possibility of having the overflow problem even when circular arrays were used to solve the overflow. However, in functional languages there are no such problems. The entire memory is under the control of the programmer. The only limitation is the size of the memory.

(2) A garbage collection mechanism is supported with the linked lists in functional language. Compared with other languages such as C and C++, modelling using CPN-ML releases the designer from any kind of memory management problems. This mechanism relieves the burden of writing additional codes to manage cells of memory that are no longer in use. Without such a mechanism there would be more overheads to the design specification.

CONCLUSIONS

In this paper, we present a new concurrent multithreaded model that extends the work-stealing scheduling technique by developing a new strategy that improves its performance. The RSS strategy provides a new means of balancing the threads among the modelled cores. It is based on balancing thread numbers in the model through moving threads from the wealthiest (victim) core to the non-working (thief) cores. The strategy has several beneficial features:

- Simplicity: Implementing the strategy is simple and does not need any complicated calculations.

• Scalability: The mechanism of the strategy has been designed to be unlimited. The model can be easily expanded to deal with any number of cores.
• Generality: This strategy can deal with any kind of divide-and-conquer problems. Although in this paper we used this strategy to balance the threads of the moves in the Tower of Hanoi game, the strategy can deal with any divide-and-conquer problem.
• Concurrency: One of the main motivations behind the development of this strategy was to make all the cores work as much as possible. This strategy is superior to the previous one, the In-Order strategy, in the increase of the number of concurrent steps among the modelled cores.

FUTURE WORK

One of the future plans in this area is the development of more work-stealing strategies. It is envisaged that these future strategies should lead to better performance especially with regard to the balancing of the threads. In addition, compared to those strategies that currently exist, these new strategies should lead to a shorter run-time.

REFERENCES

1. S. Hofmeyr, C. Iancu and F. Blagojevic´, "Load balancing on speed", Proceedings of 15th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming, **2010**, Bangalore, India, pp.147-158.
2. R. D. Blumofe and C. E. Leiserson, "Scheduling multithreaded computations by work stealing", *J. ACM*, **1999**, *46*, 720-748.
3. N. S. Arora, R. D. Blumofe and C. G. Plaxton, "Thread scheduling for multiprogrammed multiprocessors", Proceedings of 10th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA), **1998**, Puerto Vallarta, Mexico, pp.119-129.
4. D. Hendler, Y. Lev, M. Moir and N. Shavit, "A dynamic-sized nonblocking work stealing deque", *Distrib. Comput.*, **2005**, *18*, 189-207.
5. D. Hendler and N. Shavit, "Non-blocking steal-half work queues", Proceedings of 21st Annual Symposium on Principles of Distributed Computing, **2002**, Monterey, California, USA, pp.280-289.
6. U. A. Acar, G. E. Blelloch and R. D. Blumofe, "The data locality of work stealing", Proceedings of 12th Annual ACM Symposium on Parallel Algorithms and Architectures, **2000**, Bar Harbor, Maine, USA, pp.1-12.
7. D. Chase and Y. Lev, "Dynamic circular work-stealing deque", Proceedings of 17th Annual ACM Symposium on Parallelism Algorithms and Architectures, **2005**, Las Vegas, Nevada, USA, pp.21-28.
8. A. M. Al-Obaidi and S. P. Lee, "A concurrent multithreaded scheduling model for solving Fibonacci series on multicore architecture", *Int. J. Adv. Comput. Technol.*, **2011**, *3*, 24-37.
9. A. M. Al-Obaidi and S. P. Lee, "A concurrent coloured Petri nets model for solving binary search problem on a multicore architecture", Proceedings of 2nd International Conference on Software Engineering and Computer Systems, **2011**, University Malaysia Pahang, Malaysia, pp.463-477.

10. K. Jensen and L. M. Kristensen, “Coloured Petri Nets: Modelling and Validation of Concurrent Systems”, Springer, Dordrecht**, 2009**.
11. K. Jensen, “An introduction to the practical use of coloured Petri nets”, in “Lectures on Petri Nets II: Applications, Advances in Petri Nets” (Ed. W. Reisig and G. Rozenberg), Springer-Verlag, London, **1998**, pp.237-292.
12. L. M. Kristensen and S. Christensen, “Implementing coloured Petri nets using a functional programming language”, *Higher-Order Symbol. Comput.*, **2004**, *17*, 207-243.
13. N. A. Mulyar and W. M. P. van der Aalst, “Patterns in colored Petri nets” , Working Paper, **2005**, Eindhoven University of Technology, Eindhoven, Netherlands.
14. J. Reppy, “Standard ML of New Jersey (SML/NJ)”, Bell Laboratories, New Jersey, USA, http://www.smlnj.org/ (Accessed: January 2012).
15. E. R. Gansner and J. H. Reppy, “The Standard ML Basis Library”, University of Cambridge Press, Cambridge, **2002**.
16. J. D. Ullman, “Elements of ML Programming”, Prentice-Hall, Upper Saddle River, **1998**.
17. CPN Group, “CPN Tools”, Eindhoven University of Technology, Eindhoven, Netherlands, http://cpntools.org/ (Accessed: January 2012).
18. T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, “Introduction to Algorithms”, MIT Press, Cambridge, **2001**.
19. F. P. Miller and A. F. Vandome, “Divide and Conquer Algorithm”, Alphascript Publishing, Beau Bassin, **2010**.
20. K. M. Kavi, A. Moshtaghi and D. J. Chen, “Modeling multithreaded applications using Petri nets”, *Int. J. Parallel Program.*, **2002**, *30*, 353-371.
21. J. L. Peterson, "Petri nets", *ACM Comput. Surv.*, **1977**, *9*, 223-252.
22. T. Murata, "Petri nets: Properties, analysis and applications", *Proc. IEEE*, **1989**, *77*, 541-580.

Maejo Int. J. Sci. Technol. **2012**, *6*(02), 297-314

Maejo International Journal of Science and Technology

ISSN 1905-7873
Available online at www.mijst.mju.ac.th

Full Paper

Development of a simple force prediction model and consistency assessment of knee movements in ten-pin bowling

Chung-Shun Hung [1], Li-Cheng Hsieh [1,]* and Hong-Wen Wu [2]

[1] Department of Bio-Industrial Mechatronics Engineering, The College of Agriculture and Natural Resources, National Chung Hsing University, 250 KuoKuang Road, Taichung, Taiwan 402

[2] Department of Physical Education and Graduate Institute of Physical Education, National Taiwan College of Physical Education, 16 ShuangShih Road, Taichung, Taiwan 404

* Corresponding author, e-mail: lchsieh@nchu.edu.tw & lchsieh081559@gmail.com

Received: 20 May 2011 / Accepted: 28 July 2012 / Published: 7 August 2012

Abstract: The aim of this research is to use LabVIEW to help bowlers understand their joint movements, forces acting on their joints, and the consistency of their knee movements while competing in ten-pin bowling. Kinetic and kinematic data relating to the lower limbs were derived from bowlers' joint angles and the joint forces were calculated from the Euler angles using the inverse dynamics method with Newton-Euler equations. An artificial-neural-network (ANN)-based data-driven model for predicting knee forces using the Euler angles was developed. This approach allows for the collection of data in bowling alleys without the use of force plates. Correlation coefficients were computed after ANN training and all values exceeded 0.9. This result implies a strong correlation between the joint angles and forces. Furthermore, the predicted 3D forces (obtained from ANN simulations) and the measured forces (obtained from force plates via the inverse dynamics method) are strongly correlated. The agreement between the predicted and measured forces was evaluated by the coefficient of determination (R^2), which reflects the bowler's consistency and steadiness of the bowling motion at the knee. The R^2 value was beneficial in assessing the consistency of the bowling motion. An R^2 value close to 1 implies a more consistent sliding motion. This research enables the prediction of the forces on the knee during ten-pin bowling by ANN simulations using the measured knee angles. Consequently, coaches and bowlers can use the developed ANN model and the analysis module to improve bowling motion.

Keywords: ten-pin bowling, knee movement, LabVIEW, artificial neural network, physical education

---

INTRODUCTION

Sports provide significant health benefits and contribute to quality of life. Ten-pin bowling is a popular international indoor sport that requires accuracy of movement, steadiness of motion, and adequate direction and use of force. Biomechanical information, such as joint angles and forces, would be useful for bowlers and their coaches to improve bowlers' movements and poses. A bowler's posture of action can be observed and improved by studying his or her biomechanics.

Ten-pin bowling requires great skill and stable motion by the bowler. Compared to unskilled bowlers, skilled bowlers have greater mental toughness and are more confident in both performing a technique and handling the equipment. Skilled bowlers rely less on luck and undertake higher levels of planning and evaluation [1-2]. Bowlers are prone to upper limb injuries, which generally occur in the hand and fingers because holding the heavy bowling ball requires the bowler to insert the thumb and two other fingers into three holes drilled in the ball. Such injuries can potentially harm tendons and ligaments [3]. Hence, most research on ten-pin bowling in recent years has focused on the upper limbs [4-6]. While bowling, the bowler should keep the lower limbs stable to avoid injury. The ability to slide the front foot consistently enables the bowler to have a predictable stable base from which to deliver the ball more accurately [7]. Lower limb injuries are related to the bowler's gait and stance while throwing the bowling ball. An improper gait and/or stance can cause adductor muscle strains, ankle sprains, knee ligament injuries and femoral shaft fractures [8]. Investigating the motion of the lower limbs during bowling produces results that could be used to reduce injury. Kinematic information taken from ten-pin bowlers can broaden coaches' knowledge to train their athletes more effectively [9]. The kinematic and kinetic variables (i.e. angle, foot position, velocity and force) for ten-pin bowling have been observed and studied [7, 9, 10].

Through kinematics calculations, the velocities of body segments and relevant movement parameters, such as joint angle, joint force and joint moment, can be calculated. Euler's equations [11] are commonly used to obtain kinematics data. Riley and Kerrigan [12] conducted an analysis of the hip, knee and ankle joints using acceleration and angular acceleration to analyse the patients' specific impairments in the lower limb joints. Because of advancements in the integration of cameras and computers, dynamic image analysis systems are available for the analysis of sport biomechanics. Chan et al. [13] used three-dimensional (3D) high-speed capture cameras to analyse the movement of baseball players and, using kinematics and kinetics calculations, compared the players' knee loading and hip position during the performance of different actions. Chu et al. [9] profiled the techniques of elite-level ten-pin bowlers in terms of position, joint angle and velocity. Stuelcken and Sinclair [10] provided normative data on the magnitude of the ground reaction forces experienced by elite bowlers upon front foot contact while bowling. However, earlier studies offered insufficient data for the complete construction of a systematic model, which requires numerous calculations. Few researchers have computed and processed the large amounts of data involved in kinematic and kinetic calculations. Using one software package is a better way to build a systematic model containing all of the necessary functions. This approach also makes it easier to display the results (as angles, forces and moments) with illustrations and to process quantities of kinematic and kinetic data.

In this study, combined calculation programmes were employed using the data acquisition programming function in LabVIEW (National Instruments). LabVIEW is a widely used graphical programming language that facilitates communication between laboratory equipment and a personal computer. LabVIEW systems are called virtual instruments [14-15]. Engineers use this software for short-term planning, development and applications because of its good adaptability and flexibility in terms of compatible tools, test design, measurements, control systems and analysis [16]. Virtual instrumentation in LabVIEW is hierarchical and modular, making it very useful for programme design. Riemann et al. [17] used LabVIEW to convert and sample vertical ground reaction force data from a force plate and to process the signal for data analysis. Gopalai et al. [18] regenerated and displayed the motion of a lawn bowler with the aid of physically attached, tri-axial wireless accelerometers on various body segments using data collected by LabVIEW virtual instrumentation. Hon et al. [19] used a wireless inertial sensor to analyse the arm swing of a ten-pin bowler and developed an interactive graphical user interface (GUI)-based LabVIEW programme to analyse and visualise the critical biomechanical parameters of the arm swing. However, these analysis systems were not able to verify the consistency of the bowling motion and lack biomechanical analyses of the lower limbs. In this study, LabVIEW was used to design a human-machine interface capable of displaying biomechanical information describing the lower limbs during ten-pin bowling.

Biomechanical parameters associated with the lower limbs while bowling, such as angles and forces, also affect the slide of the front foot. Thus, the relationships between these angles and forces must be derived and observed. Music et al. [20} constructed motion trajectories using a three-segment dynamic human body model (shank, thigh and HAT (head-arms-trunk)). The joint forces and moments were derived using a highly coupled system of differential equations. The Jacobian matrix (J) is a commonly used tool for describing human motion in 3D space [21-23]. Li et al. [21] stated that spherical coordinates and Euler angles can represent human motion and that the Jacobian matrix function can be used to calculate the dynamic equation for human motion. These studies indicate that joint angles and forces can describe 3D human motion and that the joint forces can be calculated from the joint angles using specific functions. Thus, it is possible to find and analyse the relationships between the joint angles and forces on the lower limbs during ten-pin bowling in this study.

Employing the analysis system in a bowling alley without the use of force plates and performing repeated testing would be very difficult. Therefore, in this study, the joint forces were estimated from the joint angles using artificial neural network (ANN) simulation. These simulation methods have been employed extensively in several fields. ANNs have been successfully and widely applied in the analysis of clinic biomechanics [24]. Barton and Lees [25] used an ANN to perform an automated diagnosis of gait patterns classified by hip-knee angles. A back-propagation neural network was used to carry out biomedical signal analysis [26-27]. Barton et al. [28] used an ANN to quantify the deviation of a patients' gait from a normal gait using kinetic and kinematic data. Uchiyama et al. [29] determined the static relationships between the elbow joint angle and the joint torque using an ANN technique in which integrated electromyograms and joint angles were the inputs and the torque was the output. Thus, in this study an ANN model was constructed for the prediction and simulation of joint forces using joint angles.

To reduce the time required for ANN simulation using a large data set and to facilitate observation of the bowling motion in the lower limbs, this study focuses on the knee joint for analysis.

When a bowler slides on a surface, more weight is placed on the sliding leg. For the duration of the horizontal movement during floor contact, sliding decreases the loading of the ankle joint and increases the loading of the knee joint [30]. The knee joint has a large impact on the lower limbs in terms of reactive stability control [31]. The conditions of the bowling motion can therefore be observed most easily in the knee joint. Additionally, changes in the bowling movement result in a greater range of motion for the knee joint than for other joints. Therefore, in this study, the knee is the main source of data for the evaluation and analysis of bowlers' lower-limb motion. Accuracy and consistency have been recognised as important factors in ten-pin bowling performance [7]. Razman et al. [7] indicated that the consistency of the side-to-side foot path at the start of the slide is important for successful bowling. The objective of this research is to develop an analysis model for a simple prediction of joint forces using the measured angles of the knee without using force plates, and to assess the consistency of the knee in ten-pin bowling.

MATERIALS AND METHODS

Participants

Eight bowlers from a Taiwanese ten-pin bowling club (age: 28.8 ± 6.9 years; height: 1.71 ± 0.05 m; weight: 77.5 ± 22.1 kg) voluntarily participated in the study. None of the participants had a history of lower-limb surgery, nor had they suffered any limb injuries during the three months prior to the study. All of the participants provided written consent to the experimental protocol, which was approved by the institutional review board.

Data Collection and Processing

Six high-speed infrared cameras were used to collect the trajectories of the reflective markers. A 3D motion analysis system (Vicon MX) with two AMTI force plates (type OR6-6, 1000 Hz) was used to record the motion of and the force applied by the participants' feet. Human body characteristics (the reflective markers' positions) were used to define the coordinate positions. The angular movement, angular velocity and angular acceleration of the joints were calculated using the markers' positions. Measurements of the ground reaction forces and moments were obtained from the force plates. Each participant was involved in 10-20 trials for analysis.

The laboratory coordinate system was used as the reference to describe the 3D motion. A rotation matrix was used to describe the rotation of an object, and the relative motion and rotation matrix represent any two segments of the human body [32-34]. The Euler angles were used to analyse the angular motion of the rotating joints (Eq. (1)). Figure 1 shows the 3D coordinate system

$$R_{Y-X-Z}(\alpha,\beta,\gamma)=\begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}=\begin{bmatrix} S\alpha S\beta S\gamma + C\alpha C\gamma & S\alpha S\beta C\gamma - C\alpha S\gamma & S\alpha C\beta \\ C\beta S\gamma & C\beta C\gamma & -S\beta \\ C\alpha S\beta S\gamma - S\alpha C\gamma & C\alpha S\beta C\gamma + S\alpha S\gamma & C\alpha C\beta \end{bmatrix} \tag{1}$$

where

$\beta = \tan^{-1}\left(-R_{23}, \sqrt{R_{13}^2 + R_{33}^2}\right)$,
$\alpha = \tan^{-1}\left(R_{13}/C\beta,\ R_{33}/C\beta\right)$,
$\gamma = \tan^{-1}\left(R_{21}/C\beta, R_{22}/C\beta\right)$,
S = sine, C = cosine, and $R_{Y\text{-}X\text{-}Z}$ is a rotation matrix for two segments of human body.

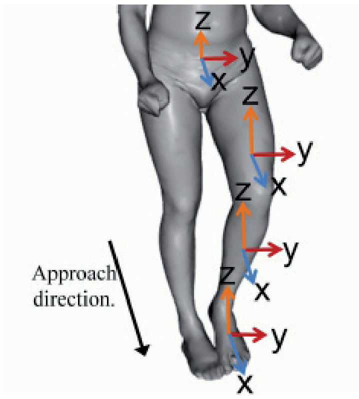


Figure 1. The coordinate system defined for the lower extremities

for every joint of the lower limb. The rotation angles about the y-axis (angle of flexion and extension), x-axis (angle of abduction and adduction) and z-axis (angle of internal and external rotations) are represented by α, β and γ respectively. The measurements taken from the plates were used to calculate the forces and moments of the joints of the lower limbs using the inverse dynamic method with the Newton-Euler equations (2 and 3) [35-37]:

$$\vec{F}_p = m\vec{a} - \vec{F}_d - m\vec{g} \tag{2}$$

$$\vec{M}_p = I\dot{\vec{\omega}} + \vec{\omega} \times I\vec{\omega} - \left(\vec{M}_d + \vec{r}_d \times \vec{F}_d + \vec{r}_p \times \vec{F}_p\right) \tag{3}$$

where m is the segment mass (kg); $\vec{a}$, the segment acceleration of the centre of mass (m/s^2); g, the acceleration due to gravity (m/s^2); $\vec{F}_p$, the proximal end joint force (N); $\vec{F}_d$, the distal end joint force (N); $\vec{M}_p$, the proximal end joint moment (N-m); $\vec{M}_d$, the distal end joint moment (N-m); I, the moment of inertia of the segment (kg/m^2); ω, the angular velocity of the segment (rad/s); $\vec{r}_d$, the distal end moment arm (m); and $\vec{r}_p$, the proximal end moment arm (m).

Derivation of Functions for Angles and Forces

The experimental data were collected from the bowlers in the laboratory. The bowling motion can be analysed using sport biomechanics, and the inverse dynamics method can be used to calculate the joint forces and moments from the ground reaction forces measured by force plates. However, force plates cannot be used in bowling alleys or during competitions. Instead, re-evaluating and predicting the forces or moments using an experimental database built from laboratory data is preferred. Thus, the Jacobian matrix function can represent the motion in ten-pin bowling as shown in Eq. (4) - Eq. (7) [21]:

$$p_i(x,y,z) = \begin{bmatrix} f_1(\phi,\varphi) \\ f_2(\phi,\varphi) \\ f_3(\phi) \end{bmatrix}, \tag{4}$$

where p_i is the position vector of joint i, and φ and ϕ are compressed parameters representing the three Euler angles. In the spherical coordinate system, the matrix is expressed as

$$f_1(\phi,\varphi) = \sum_{j=1}^{n} r_j \cdot \sin(\phi_j)\cos(\varphi_j),$$

$$f_2(\phi,\varphi) = \sum_{j=1}^{n} r_j \cdot \sin(\phi_j)\cos(\varphi_j),$$

$$f_3(\phi) = \sum_{j=1}^{n} r_j \cos(\phi_j),$$

where r_j is the length of segment j.

$$\dot{p}_n = J\dot{\theta}_n \tag{5}$$

where θ_n is composed of all angles (from α, β and γ) for each joint, and

$$J = \begin{bmatrix} \frac{\partial f_1(\phi,\varphi)}{\partial \phi_j} \, \frac{\partial f_1(\phi,\varphi)}{\partial \varphi_j}, \frac{\partial f_2(\phi,\varphi)}{\partial \phi_j} \, \frac{\partial f_2(\phi,\varphi)}{\partial \varphi_j}, \frac{\partial f_3(\phi)}{\partial \phi_j} \; 0 \end{bmatrix} \tag{6}$$

Using Newton's laws of motion, the equation can be expressed as

$$F = ma = m \cdot \frac{\partial v}{\partial t} = m \cdot \frac{\partial^2 P(x,y,z)}{\partial t^2} = m \cdot \frac{\partial (J\dot{\theta})}{\partial t}. \tag{7}$$

If the mass and length of the segments were to remain constant, the forces would be relative to only the angles. Thus, the relationships between the joint angles and forces on the lower limbs for bowling might be correlated. However, before the joint angles and forces can be analysed, the initial calculations require a computer software programme to process the large amount of data and to calculate the kinematics and kinetics. The functions for these calculations were combined using LabVIEW, and human-machine interface graphics were displayed using a multi-programme analysis. This approach is helpful for observing the phenomenon and the results of the ten-pin bowling motion.

Development of Graphical User Interface (GUI) Modular System

LabVIEW virtual instrumentation is useful in designing and planning an operating interface to analyse human motion in terms of kinematics and kinetics. In this study, the GUI modular analysis system for the lower limbs of bowlers engaged in ten-pin bowling was designed using LabVIEW (Figures 2 and 3). To import human parameters and experimental data, biomechanical descriptors such as joint angles and forces were calculated and displayed in the GUI panel. The biomechanical information describing each bowler was exported, saved and analysed using the system.

The developed GUI panel consists of three tabs: the bowler's information, motion analysis, and ANN analysis and application. In Figure 2, the bowler's information tab shows the bowler's basic information, usual bowling hand and support foot (whether the final step is taken with the left or right foot). The bowler's basic information includes sex, age, height, weight and years of experience in bowling. In the drawing in the middle of Figure 2, the bowler's usual bowling hand and support foot are shown in colour (right: red, left: blue).

The motion analysis tab, shown in Figure 3, imports the Vicon data (the bowler's height and weight), and the analysis system performs computations in terms of the kinematics and kinetics. The left side of the tab shows the original data acquired from the two force plates, the joint angles of the lower limbs (hip, knee and ankle), the joint forces and the joint moments. The right side of the tab displays the 3D dynamic motion of the lower limbs. The orientation of the 3D picture is adjustable, facilitating observation of the changes in the joint angles or motions.

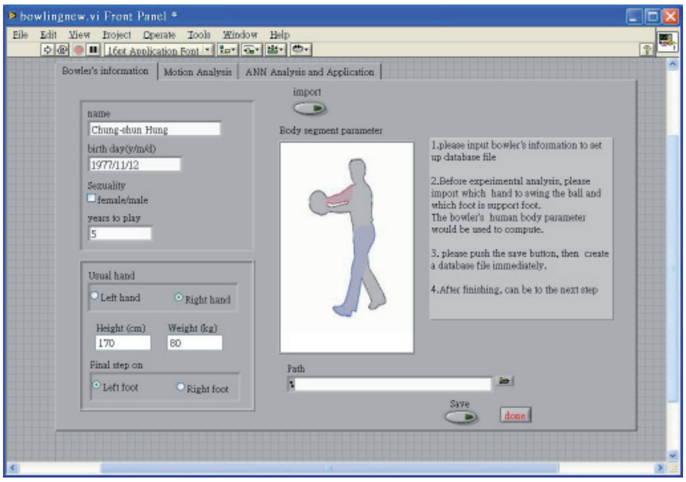


Figure 2. Bowler's information tab. The bowler's usual bowling hand and support foot are shown in colour (right: red, left: blue).

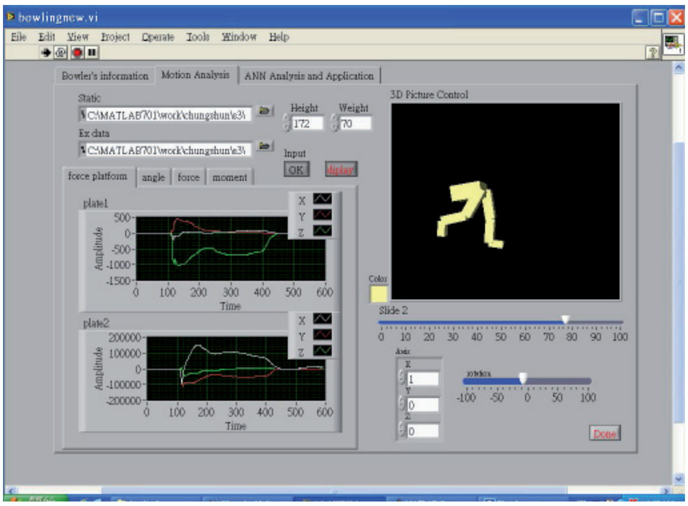


Figure 3. Motion analysis tab

In the ANN analysis and application tab, the joint angles and forces of each bowler's knee were exported from the motion analysis tab and simulated using an ANN model. To confirm and interpret the forces predicted from joint angles during the dynamic bowling motion, the system in this tab consists of LabVIEW with a developed three-layer network ANN model. Thus, coaches and bowlers can apply the ANN model for simple force prediction and assessment of knee position consistency during bowling.

ANN Model

The present study focused on the 3D Euler angles of the lower limb joints. Without force plates, the joint forces were estimated from the Euler angles. While bowling, the knee joint is likely to undergo motion. The accuracy and consistency of knee motion has a large impact on the lower limbs during bowling. An ANN model was constructed for each bowler, modelling the joints of the lower limbs based on the experimental data obtained in this study. The main aim of the study was to conduct trials and verify the simulated results relating to the movement of the knee joint, with a focus on performing a simple assessment of the knee-joint movements performed by bowlers and the consistency of those movements.

The joint forces are related to variations in the Euler angles through the transfer function (Eq. (7)) derived from the Jacobian matrix and differential equations. To enhance each bowler's unique transfer function, an ANN function capable of learning and training was created to replace the function in Eq. (7). The ANN architecture was constructed with one input layer, one hidden layer and one output layer to create a three-layer neural network. The ANN model of the knee consists of three neurons in the input layer, five neurons in the hidden layer, and three neurons in the output layer (Figure 4(a)). The Levenberg-Marquardt back propagation algorithm [38] was used for ANN training. A tangent sigmoid function was employed in both hidden layers and a linear transfer function was applied to the output layer. The model was trained using back propagation to predict the dynamic joint forces of the knee using the 3D Euler angles as the input data. Each participant was involved in 10-20 trials, each of which contributed 100-200 samples; thus, more than 1000 knee angles were available for use as input data for each bowler's ANN model. The ANN model was trained with back propagation using the opposing forces to the 3D joint forces of the knee (Fx, Fy and Fz versus the x-axis, y-axis and z-axis of the knee respectively). The ANN model would then be able to predict joint forces experienced in bowling alleys or in other situations without the use of force plates (Figure 4(b)).

Data Analysis and Consistency Assessment

The trend line for the relationship between the knee angles and the joint forces is found using linear regression on the sagittal plane. The weight and bias parameters for the network can be found after training the ANN model, and experimental data can be substituted into the network to simulate new results. If the correlation coefficient between the original measured forces and the simulated outputs is greater than 0.9 [39], then the transfer function relating joint angle and force in the ANN model can be built. The coefficient of determination (R^2) reflects the degree of fit or error for the ANN model of the knee while bowling and indicates the degree of convergence and accuracy of the simulation compared to the experimental data. The closeness of the R^2 value to 1 indicates the degree of success of the ANN model's simulation [40-41]. If the R^2 value is large, then the data used for training the ANN model converge easily. The R^2 value also reflects the precision of the experimental data and can therefore represent the consistency of the knee's bowling motion. If the R^2 value is small, then the data are divergent and the variation in the bowler's knee motion is large. Thus, the consistency and steadiness of a bowler's knee joints can be evaluated on the basis of the R^2 value, which is larger for bowlers with a higher degree of consistency.

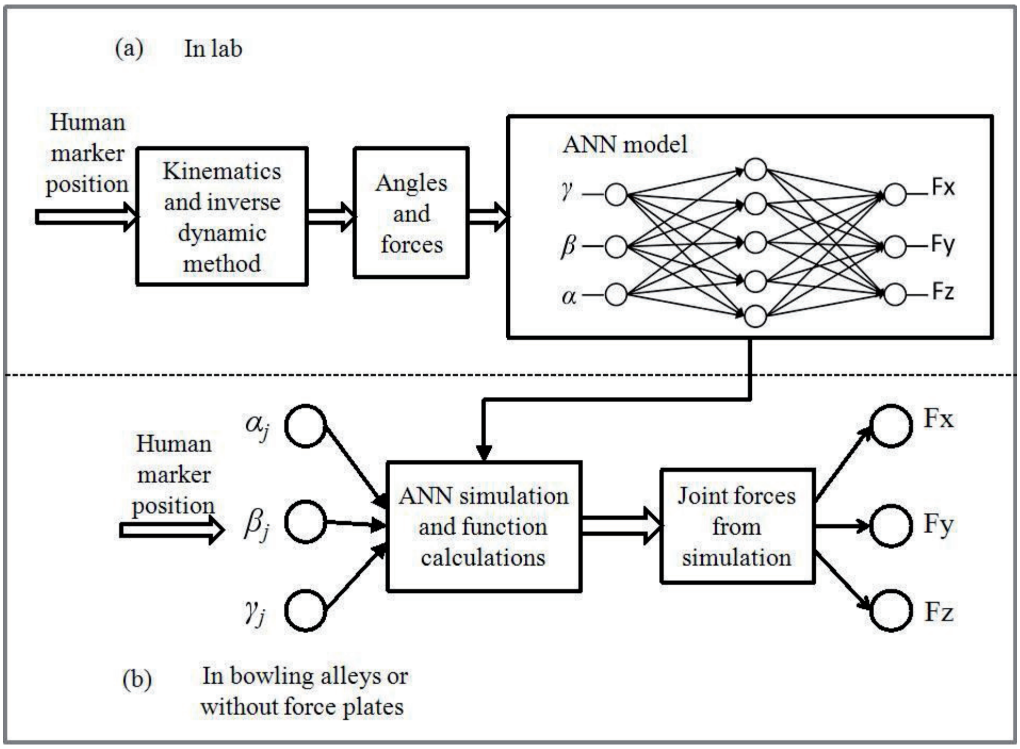


Figure 4. Structure of ANN model for prediction and simulation: (a) in laboratory – deriving the forces and angles used in ANN model from experimental data (α is rotation angle of flexion and extension versus Fz; β is rotation angle of abduction and adduction versus Fy; and γ is angle of internal and external rotation versus Fx); (b) in bowling alleys or without force plates – using ANN model obtained in laboratory with measured angles (α_j, β_j and γ_j) to predict joint forces.

RESULTS

Data on Angles and Forces

Bowlers need to possess great strength and speed to perform well in ten-pin bowling, but over-pursuing these requirements usually results in undetectable limb injuries. In the final step of the ten-pin bowling motion (Figure 5), the leg exerts extra force to allow the bowler to slide, thereby causing the bowler's body weight to be thrust in the forward direction.

The force plate signals were recorded starting when the support foot touches the plate and stopping at the end of the sliding motion (Figure 6(a)). The force plate data obtained from the bowling dynamic motion was used to analyse the sources of force. Then the data computed in terms of kinematics and kinetics was processed over the same range (e.g. the intercepting range of knee forces and angles shown in Figure 6). In Figure 6, the Fz on the knee decreased during the period ranging from touching the plate to stopping the slide.

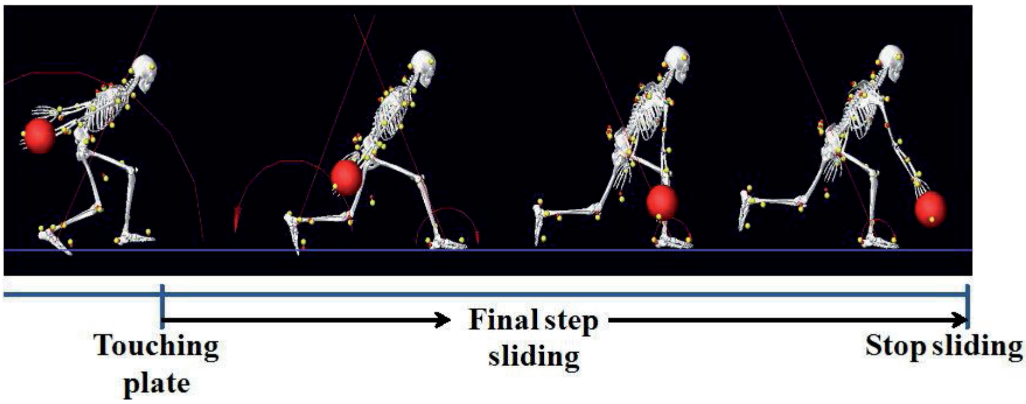
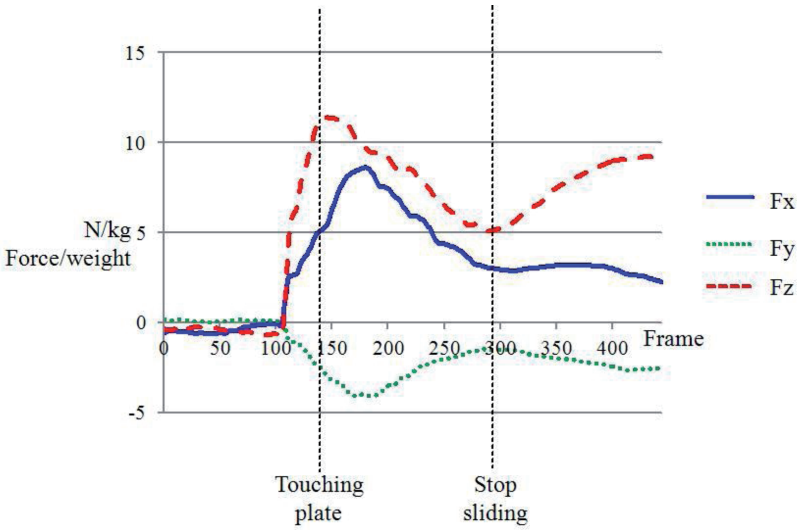
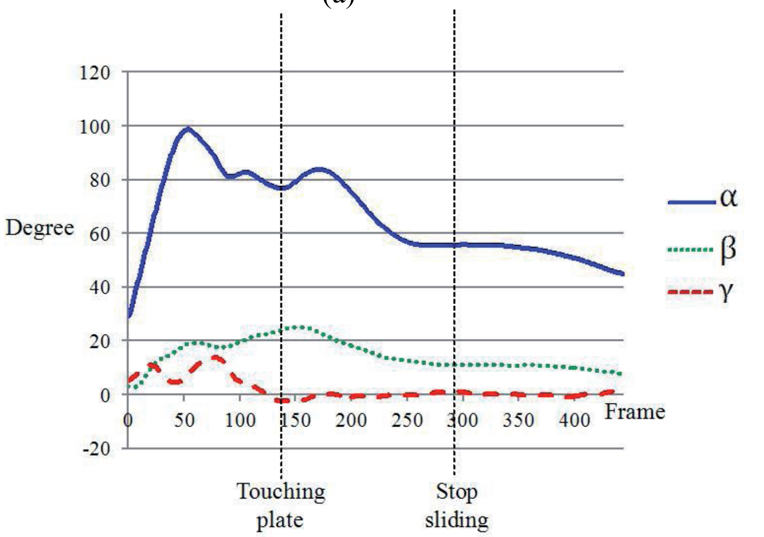


Figure 5. Illustration of the motions used during ten-pin bowling, created using LifeMOD™ (based on MSC.ADAMS®, Mechanical Dynamics Inc., USA) simulation



(a)



(b)

Figure 6. Determination of the fitted data range for the forces (a) and angles (b) of the knee joint

The angles and forces obtained from the processed data were exported to Microsoft Excel to construct distribution diagrams. Figure 7(a) shows the distribution of the knee angles during flexion under forces in the sagittal plane. The x-axis shows the knee angles (°), and the y-axis shows the vertical forces of the knee (N/kg). Figure 7(b) shows the distribution of data and the trend line obtained by a simple linear regression. The mean trend line slope for the eight bowlers is 0.14 N/kg-degree, which indicates that the vertical forces are positively correlated with the joint angles.

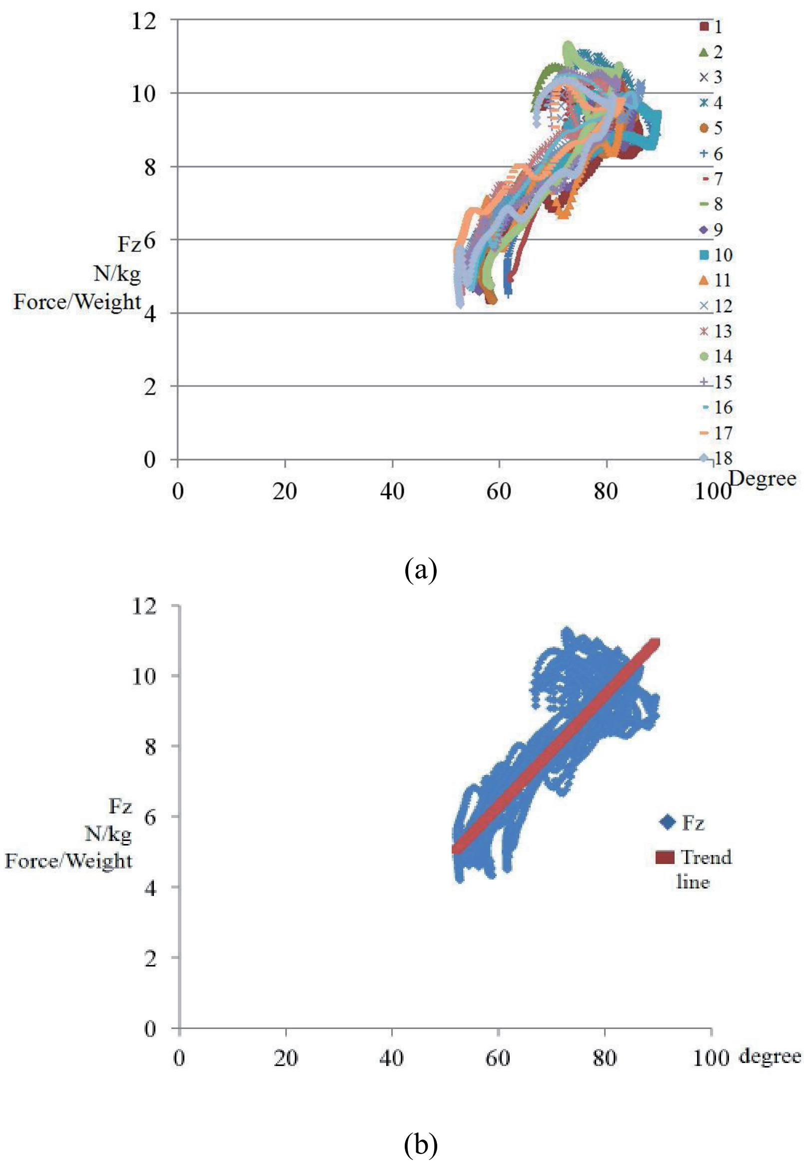


Figure 7. (a) Distribution of knee angles and forces during flexion (trial numbers: 1-18); (b) Regression of knee angles and forces during flexion

Analysis Using ANN Model

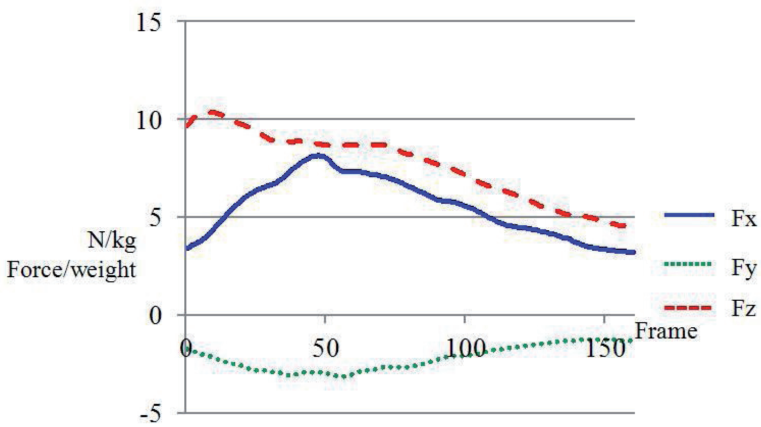
After the ANN model training, the correlation coefficients between the joint angles and forces were determined. The correlation coefficients between the original and simulated knee forces during

3D dynamic motion of eight bowlers obtained by the ANN model training are listed in Table 1. This approach can therefore be used to predict the forces on the knee joint of a competing bowler in competition. Figure 8 shows one of the subjects' changes in knee forces as predicted by the knee angles using the ANN model.

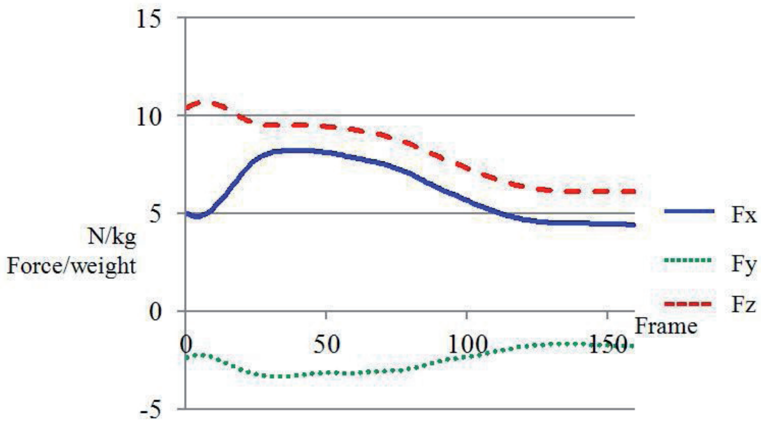
Table 1. Correlation coefficients between original and simulated knee forces for eight bowlers as determined by training the ANN model

r	S1	S2	S3	S4	S5	S6	S7	S8	Mean
r1	0.86	0.86	0.94	0.92	0.95	0.93	0.98	0.98	0.93
r2	0.93	0.88	0.91	0.81	0.92	0.87	0.96	0.91	0.90
r3	0.86	0.91	0.96	0.84	0.86	0.95	0.96	0.94	0.91

Note: r = correlation coefficient between original and simulated forces;
r1, r2 and r3 = correlation coefficients for Fx, Fy and Fz respectively;
S1-S8 denote bowlers number 1-8.



(a)



(b)

Figure 8. (a) Forces on the knee joint acquired from experimental data obtained with force plates, as determined by the inverse dynamics method; (b) The knee forces predicted from the knee angles using ANN simulation model.

The forces predicted by the model were compared to the original forces for the new data sets to determine whether the R^2 values agreed [42]. Table 2 lists the R^2 values for the knee consistency assessment of the eight participants in the study.

Table 2. Validation of ANN model and consistency assessment using ANN

R^2	S1	S2	S3	S4	S5	S6	S7	S8
$(R1)^2$	0.86	0.73	0.89	0.84	0.89	0.87	0.97	0.96
$(R2)^2$	0.86	0.77	0.83	0.65	0.85	0.76	0.93	0.83
$(R3)^2$	0.74	0.82	0.92	0.71	0.74	0.91	0.92	0.88

Note: The coefficient of determination (R^2) reflects the degree of fit between the ANN model and the experimental data. $(R1)^2$, $(R2)^2$, and $(R3)^2$ were used to assess the consistency of Fx, Fy and Fz respectively. S1-S8 denote bowlers number 1-8.

DISCUSSION

Of the knee forces along the x-, y-, and z-axes, the vertical force Fz is greater than the forces in the other directions (shown in Figure 6). At this moment, a large amount of weight is placed on the sliding leg, which must be able to withstand the majority of the bowler's body weight and bear the force loading and the speed. A high lateral force Fy would adversely affect the consistency of the knee during bowling. The change in the range of motion of the ankle joint is relatively small, but this change imposes a burden on the knee in flexion and causes pain. An increase in the burden on the knee joint may result in injuries and inflammation. Patellar tendinitis of the knee is often induced by inflammation of the tendon between the patella and the shinbone. Bowling on a wooden sliding surface during floor contact may decrease the loading of the ankle joint but increase the load on the knee joint [30]. During the slide, the angle α of the knee decreased and the knee extended during the motion from touching the plate to stopping the slide (Figure 6). In similar sports such as cricket bowling, the vertical ground reaction forces are greater than the horizontal forces and fast bowlers use an extended knee during the front-foot contact phase [10, 43]. Therefore, a bowler must pay attention to his or her knee motion while moving the lower limbs during ten-pin bowling.

As the bowlers' legs slided on the plate, the range of knee angles was from 50° to 90° for flexion in the sagittal plane and the range of the vertical force Fz was approximately 4-12 N/kg (Figure 7). Linear regression was used to analyse the distribution of the angles and forces, and the mean slope of the trend line was 0.14 (Figure 7(b)). Thus, as the knee angles increased during flexion while bowling, the joint forces increased as well. Suter and Herzog [44], through EMG measurements, noticed that the muscle torque increased with increasing knee angles. Smith et al. [45] determined the joint angles and axial-joint contact forces during stair climbing and squatting activities. They found that an increase in the knee angles during flexion in the sagittal plane caused an increase in the axial tibio-femoral joint contact forces. These two studies illustrate that the forces on the knee joint increase with increasing knee angle. The results of our study agree with these previously reported results.

The motion of the bowler's knee was described by the individual transfer function obtained from the ANN model. Table 1 lists the correlation coefficients between the experimental and ANN model simulated knee forces. The means of the correlation coefficients for the eight bowlers are: r1 = 0.93, r2 = 0.9 and r3 = 0.91 for Fx, Fy and Fz respectively. All the means of the correlation coefficients are greater than 0.9. Additionally, the correlation coefficient for bowler 7 (S7) is the highest. On the basis of this validation of the ANN model, the joint forces can be predicted and the angles and forces can be interpreted as having a high correlation with and influence on the consistency of the knee. Force plates cannot be used in bowling alleys or during competitions; thus, the use of the ANN model is a good option. Changes in motion can be determined from the 3D knee forces predicted from the knee angle inputs into the ANN simulation model. Figure 8 shows one of the subjects' knee forces predicted using the ANN simulation model with the knee angles as input. The knee forces computed from the measured data and the 3D knee forces predicted by the ANN model are in good agreement. Therefore, the knee forces of the sliding leg during a ten-pin bowling competition can be predicted without the use of force plates.

In Table 2, a higher R^2 value implies greater accuracy and consistency of a bowler's knee joint. The R^2 values for the knee joint of bowler 7 (S7) from the ANN model are 0.97 in Fx, 0.93 in Fy and 0.91 in Fz. The consistency of the knee movement of bowler 7 is greater than that of the other bowlers. The R^2 values of bowler 4 (S4) are 0.84 in Fx, 0.65 in Fy and 0.71 in Fz, indicating that the knee's consistency is low and that the experimental data for the bowler are divergent. The R^2 values for the lateral and vertical forces for bowler 4 (S4) are even lower, implying unsteadiness, and these low values affect knee abduction/adduction and flexion/extension respectively.

The accuracy and consistency of knee movement is an important factor for assessing bowling performance and conditions. The R^2 value, which is calculated from the results of the ANN model, indicates the degree of convergence and accuracy of the simulation in comparison to experimental data. Because force plates cannot be used in bowling alleys or competitions, the R^2 value is an effective metric for assessing the level of consistency and steadiness of the knee during ten-pin bowling. This value provides bowlers and coaches with useful information for observing and analysing the bowling motion.

LabVIEW was used to construct a model for simple force prediction and consistency assessment of the knee during ten-pin bowling. The model was developed to collect and process data, filter signals and perform computations of the kinetics and kinematics. The angles, forces and moments of the lower-limb joints (hip, knee and ankle) could also be viewed with corresponding diagrams. The developed analysis module is simple to operate and gives bowlers the opportunity to view their joint conditions and motions during bowling. An individual ANN model incorporated into the LabVIEW model could be used to predict knee forces using measured knee angles for each bowler without the use of force plates. Coaches and bowlers can thus assess the consistency of knee motion with the R^2 values displayed in the analysis module on the GUI panel and can review and analyse the bowling motion to enhance performance.

CONCLUSIONS

An analysis module and software with a GUI interface in LabVIEW for consistency assessment of the knee during bowling has been developed. The module processes signals from high-speed cameras and force plates, computes the kinetics and kinematics, generates a 3D diagram of the lower limbs, executes the learning and simulation algorithms for the ANN model, and performs a consistency assessment. The developed module is suitable for use in bowling alleys or during competitions where force plates are not practical. It is useful and easy to operate, enabling bowlers to study their knee motions while bowling and, consequently, to improve their consistency for better bowling performance.

ACKNOWLEDGEMENTS

Thanks are due to all members of the Computer Aided Design Lab, National Chung Hsing University, Taiwan, for their valuable efforts to support this research. In addition, without the help of the eight bowlers from one of the Taiwanese ten-pin bowling clubs, this research would not have been possible.

REFERENCES

1. P. R. Thomas, P. J. Schlinker and R. Over, "Psychological and psychomotor skills associated with prowess at ten-pin bowling", *J. Sports Sci.*, **1996**, *14*, 255-268.
2. A. E. DeVan and H. Tanaka, "Declines in ten-pin bowling performance with advancing age", *Age Age.*, **2007**, *36*, 693-694.
3. N. Barton, "Sports injuries of the hand and wrist", *Br. J. Sports Med.*, **1997**, *31*, 191-196.
4. B. Tan, A. R. Aziz and T. K. Chuan, "Correlations between physiological parameters and performance in elite ten-pin bowlers", *J. Sci. Med. Sport*, **2000**, *3*, 176-185.
5. B. Tan, A. R. Aziz, K. C. Teh and H. C. Lee, "Grip strength measurement in competitive ten-pin bowlers", *J. Sports Med. Phys. Fitness*, **2001**, *41*, 68-72.
6. F. K. Fuss, "Design of an instrumented bowling ball and its application to performance analysis in tenpin bowling", *Sports Technol.*, **2009**, *2*, 97-110.
7. R. Razman, W. Abas and N. Othman, "Front foot slide variability and its relation to tenpin bowling performance", Proceedings of 28th International Conference on Biomechanics in Sports, **2010**, Marquette, MI, USA, pp.1-2.
8. K. C. Hsiao, M. C. Chen and C. H. Tu, "Bowling injuries", *J. Orthop. Surg. Taiwan*, **1996**, *13*, 111-114.
9. D. P. K. Chu, B. M. Zhang and K. Mau, "Tenpin bowling technique on elite player", Proceedings of 20th International Symposium on Biomechanics in Sports, **2002**, Cáceres, Spain, pp.123-125.
10. M. C. Stuelcken and P. J. Sinclair, "A pilot study of the front foot ground reaction forces in elite female fast bowlers", *J. Sci. Med. Sport*, **2009**, *12*, 258-261.
11. R. Dumas, E. Nicol and L. Chèze, "Influence of the 3D inverse dynamic method on the joint forces and moments during gait", *J. Biomech. Eng.*, **2007**, *129*, 786-790.

12. P. O. Riley and D. C. Kerrigan, "Kinetics of stiff-legged gait: Induced acceleration analysis", *IEEE Trans. Rehabil. Eng.*, **1999**, *7*, 420-426.
13. M. S. Chan, C. F. Huang, J. H. Chang and T. W. Kernozek, "Kinematics and kinetics of knee and hip position of female basketball players during side-step cutting with and without dribbling", *J. Med. Biol. Eng.*, **2009**, *29*, 178-183.
14. N. Kehtarnavaz, N. Kim and I. Panahi, "Digital signal processing system design: Using LabVIEW and TMS320C6000", Proceedings of 11th Digital Signal Processing Workshop and 3rd IEEE Signal Processing Education Workshop, **2004**, Taos Ski Valley, NM, USA, pp.10-14.
15. N. Kehtarnavaz and C. Gope, "DSP system design using Labview and Simulink: A comparative evaluation. acoustics", Proceedings of 2006 IEEE International Conference on Acoustics, Speech and Signal Processing, **2006**, Toulouse, France, pp.II-985 - II-988.
16. C. M. Lee, Y. L Liu, H. W. Shieh and C. C. Tong, "LabVIEW implementation of an auto-tuning PID regulator via Grey-predictor", Proceedings of 2006 IEEE Conference on Cybernetics and Intelligent Systems, **2006**, Bangkok, Thailand, pp.1-5.
17. B. L. Riemann, R. J. Schmitz, M. Gale and S. T. McCaw, "Effect of ankle taping and bracing on vertical ground reaction forces during drop landings before and after treadmill jogging", *J. Orthop. Sports Phys. Ther.*, **2002**, *32*, 628-635.
18. A. A. Gopalai, D. Gouwanda and S. M. N. A. Senanayake, "Human motion regeneration in 2-dimension as stick figure animation with accelerometers", *Int. J. Intell. Syst. Technol. Appl.*, **2008**, *3*, 138-148.
19. T. M. Hon, S. M. N. A. Senanayake and N. Flyger, "Biomechanical analysis of 10-pin bowling using wireless inertial sensor", Proceedings of 2009 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, **2009**, Singapore, pp.1130-1135.
20. J. Music´, R. Kamnik and M. Munih, "Model based inertial sensing of human body motion kinematics in sit-to-stand movement", *Simul. Model. Pract. Theory*, **2008**, *16*, 933-944.
21. S. Li, M. Okuda and S. I. Takahashi, "Compression of human motion animation using the reduction of interjoint correlation", *EURASIP J. Image Video Process.*, **2008**, Art. ID 693427.
22. B. Stanczyk and M. Buss, "Development of a telerobotic system for exploration of hazardous environments", Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, **2004**, Sendai, Japan, pp.2532-2537.
23. M. Meredith and S. Maddock, "Adapting motion capture data using weighted real-time inverse kinematics", *ACM Comp. Entertain.*, **2005**, *3*, 1-15.
24. W. I. Schöllhorn, "Applications of artificial neural nets in clinical biomechanics", *Clin. Biomech.*, **2004**, *19*, 876-898.
25. J. G. Barton and A. Lees, "An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams", *Gait Post.*, **1997**, *5*, 28-33.
26. M. Yuchi and J. Jo, "Heart rate prediction based on physical activity using feedforward neural network", Proceedings of 3rd International Conference on Convergence and Hybrid Information Technology, **2008**, Daejeon, Korea, pp.344-350.
27. B. Pandey and R. B. Mishra, "An integrated intelligent computing model for the interpretation of EMG based neuromuscular diseases", *Expert Syst. Appl.*, **2009**, *36*, 9201-9213.

28. G. Barton, P. Lisboa, A. Lees and S. Attfield, "Gait quality assessment using self-organising artificial neural networks", *Gait Post.*, **2007**, *25*, 374-379.
29. T. Uchiyama, T. Bessho and K. Akazawa, "Static torque-angle relation of human elbow joint estimated with artificial neural network technique", *J. Biomech.*, **1998**, *31*, 545-554.
30. B. M. Nigg, D. J. Stefanyshyn, A. I. Rozitis and A. Mündermann, "Resultant knee joint moments for lateral movement tasks on sliding and nonsliding sport surfaces", *J. Sports Sci.*, **2009**, *27*, 427-435.
31. F. Yang and Y. C. Pai, "Role of individual lower limb joints in reactive stability control following a novel slip in gait", *J. Biomech.*, **2010**, *43*, 397-404.
32. K. N. An, B. F. Morrey and E. Y. Chao, "Carrying angle of human elbow joint", *J. Orthop. Res.*, **1984**, *1*, 369-378.
33. K. N. An, A. O. Browne, S. Korinek, S. Tananka and B. F. Morrey, "Three-dimensional kinematics of glenohumeral elevation", *J. Orthop. Res.*, **1991**, *9*, 143-149.
34. E. J. Haug, "Intermediate Dynamics", Prentice-Hall, Englewood Cliffs, **1992**.
35. H. J. Woltring, "A Fortran package for generalized, cross-validatory spline something and differentiation", *Adv. Eng. Software*, **1986**, *8*, 104-113.
36. J. T. McConville, T. D. Churchill, I. Kaleps, C. E. Clauser and J. Cuzzi, "Anthropometric relationships of body and body segment moment of inertia", Technical Report AFAMRL-TR-80-119, **1980**, Aerospace Medical Research Laboratory, Wright-Pattreson Air Force Base, Dayton, Ohio, USA.
37. J. Butler, "Basic Biomechanics of the Musculoskeletal System", 3rd Edn., Lippincott Williams & Wilkins, Hagerstown (MD), **2001**, pp.176-201.
38. B. M. Wilamowski, Y. Chen and A. Malinowski, "Efficient algorithm for training neural networks with one hidden layer", Proceedings of International Joint Conference on Neural Networks, **1999**, Washington, DC, USA, pp.1725-1728.
39. M. M. Liu, W. Herzog and H. H. C. M. Savelberg, "Dynamic muscle force predictions from EMG: An artificial neural network approach", *J. Electromyogr. Kinesiol.*, **1999**, *9*, 391-400.
40. H. N. Sin, S. Yusof, N. S. A. Hamid and R. Abd. Rahman, "Optimization of enzymatic clarification of sapodilla juice using response surface methodology", *J. Food Eng.*, **2006**, *73*, 313-319.
41. M. G. Moghaddam, F. B. H. Ahmad, M. Basri and M. B. A. Rahman, "Artificial neural network modeling studies to predict the yield of enzymatic synthesis of betulinic acid ester", *Electron. J. Biotechnol.*, **2010**, *13*, 1-12.
42. S. Rath, A. P. Singh, U. Bhaskar, B. Krishna, B. K. Santra, D. Rai and N. Neogi, "Artificial neural network modeling for prediction of roll force during plate rolling process", *Mater. Manufac. Processes*, **2010**, *25*, 149-153.
43. M. Portus, B. R. Mason, B. C. Elliott, M. C. Pfitzner and R. P. Done, "Technique factors related to ball release speed and trunk injuries in high performance cricket fast bowlers", *Sports Biomech.*, **2004**, *3*, 263-284.
44. E. Suter and W. Herzog, "Extent of muscle inhibition as a function of knee angle", *J. Electromyogr. Kinesiol.*, **1997**, *7*, 123-130.

45. S. M. Smith, R. A. Cockburn, A. Hemmerich, R. M. Li and U. P. Wyss, "Tibiofemoral joint contact forces and knee kinematics during squatting", *Gait Post.*, **2008**, *27*, 376-386.

Maejo Int. J. Sci. Technol. **2012**, *6*(02), 315-322

Maejo International Journal of Science and Technology

ISSN 1905-7873
Available online at www.mijst.mju.ac.th

Short Communication

Solubility, viscosity and rheological properties of water-soluble chitosan derivatives

Siriwat Tungtong [1], Siriporn Okonogi [1], Sombat Chowwanapoonpohn [1], Weerachai Phutdhawong [2] and Songwut Yotsawimonwat [1,*]

[1] Department of Pharmaceutical Sciences, Faculty of Pharmacy, Chiang Mai University, Chiang Mai 50200, Thailand

[2] Department of Chemistry, Faculty of Liberal Arts and Science, Kasetsart University, Kamphaeng Saen Campus, Nakornpratom 73140, Thailand

* Corresponding author. e-mail: songwut_y@yahoo.com

Received: 10 June 2011 / Accepted: 6 August 2012 / Published: 17 August 2012

Abstract: An investigation and comparison of solubility, viscosity and rheological properties under neutral, acidic and alkaline conditions of water-soluble chitosan derivatives, viz. *O*-carboxymethyl chitosan, *N*,*O*-carboxymethyl chitosan, *N*-[(2-hydroxy-3-trimethylammonium)propyl] chitosan chloride and *O*-carboxymethyl-*N*-[(2-hydroxy-3-trimethylammonium)propyl] chitosan chloride, was undertaken.

Keywords: chitosan, chitosan derivatives, O-carboxymethyl chitosan, N,O-carboxymethyl chitosan, *N*-[(2-hydroxy-3-trimethylammonium)propyl] chitosan chloride, quaternised carboxymethyl chitosan, *O*-carboxymethyl-*N*-[(2-hydroxy-3-trimethyl ammonium) propyl] chitosan chloride, rheological properties

---

INTRODUCTION

Chitosan (CS) is a polysaccharide similar in structure to cellulose. It is a polycationic copolymer consisting of glucosamine and N-acetylglucosamine units. CS, commonly obtained by partial deacetylation of chitin derived from the exoskeleton of crustaceans [1], exhibits various useful physico-chemical properties such as film forming ability [2], gelation characteristics [3] and bioadhesion properties [4]. Moreover, it is biodegradable and biocompatible with low toxicity [5]. For these reasons, it has received considerable attention as a potential pharmaceutical and cosmetic excipient.

However, although CS is a very useful polymer with low toxicity, its limited solubility particularly at a physiological pH is a major obstacle for pharmaceutical and cosmetic applications. CS is a weak base with a pK_a value of 6.2-7.0 [6] due to the D-glucosamine residue, which leads to its insolubility at neutral and alkaline pH. CS dissolves in water at pH lower than 6.5, at which a substantial fraction of its amino groups are ionised. It is generally soluble in acidic solutions such as those of acetic acid, lactic acid and dilute hydrochloric acid [7]. Various chemical modifications have been used to improve its solubility at neutral and alkaline pH. CS has three reactive groups, i.e. a primary hydroxyl group at C-6 and a secondary hydroxyl group at C-3 on each repeated unit, and an amino group at C-2 on each deacetylated unit (Figure1). These reactive groups have been subjected to chemical modifications, for example reaction with glycidyl trimethylammonium chloride [8-10], carboxymethylation [11-12] and sulphonation [13-14]. The results of many studies have demonstrated that these chemical modifications can improve the solubility of CS in water over a wide range of pH. However, the rheological properties of the chemically modified CSs have not yet been fully reported. Thus, in this study we undertook to prepare four kinds of water-soluble CS derivatives, namely *O*-carboxymethyl chitosan (*O*-CMC), *N*,*O*-carboxymethyl chitosan (*N*,*O*-CMC), *N*-[(2-hydroxy-3-trimethylammonium)propyl] chitosan chloride (HTPC) and a carboxymethyl derivative of HTPC (*O*-CM-HTPC), using a specific kind of CS with a specified molecular weight and degree of deacetylation. Subsequently, the rheological properties of the water-soluble CS derivatives in aqueous solutions at different pH were investigated.

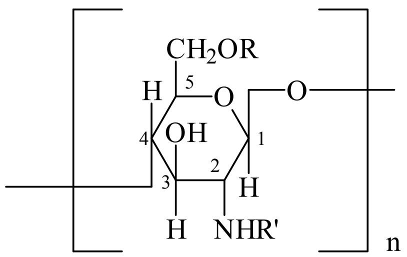


Figure 1. Structures of CS and CS derivatives: CS (R, R' = H); *O*-CMC (R = CH_2COOH, R' = H); *N*,*O*-CMC (R, R' = CH_2COOH); HTPC (R=H, R' = $CH_2CH(OH)CH_2N^+(CH_3)_3\ Cl^-$); *O*-CM-HTPC (R = CH_2COOH, R' = $CH_2CH(OH)CH_2N^+(CH_3)_3\ Cl^-$)

PREPARATION OF CS DERIVATIVES AND DETERMINATION OF DEGREE OF SUBSTITUTION

Shrimp-shell-derived CS, purchased from Seafreshchitosan Co. Ltd (Thailand), was used as starting material for the preparation of CS derivatives. Its molecular weight was about 350 kDa and the degree of deacetylation was about 84% according to the company's product information. All other chemicals used were of analytical grade.

O-CMC was prepared according to the method of Zhu et al [12]. The preparation of *N*,*O*-CMC was carried out according to the procedure reported by Hayes [15]. HTPC was prepared by a simple method as described by Spinelli et al [9], and *O*-CM-HTPC was prepared by the method as described by Sun et al [10]. All products were characterised by Fourier-transform infrared (FTIR) spectra and ^{1}H-nuclear magnetic resonance (^{1}H-NMR) spectra. The FTIR spectra of *O*-CMC, *N*, *O*-CMC, HTPC and *O*-CM-HTPC matched those reported by Sun et al [10]. The ^{1}H-NMR spectra of all products were identical to those found in the literature [8, 10, 16].

The degree of carboxymethyl group substitution (DS_{CM}) in *O*-CMC and *N,O*-CMC was determined by potentiometric titration using a calomel electrode as reference electrode and a glass electrode for the measurement, as described by Ge and Luo [11]. A sample (0.2 g) was dissolved in distilled water (40 ml) and the solution was adjusted to pH < 2 by adding 0.1N HCl. Then the solution was titrated with 0.1N NaOH and the pH value of the solution was simultaneously recorded. The amount of aqueous NaOH consumed was determined by second-order differential method. The degree of substitution was calculated as follows:

$$\mathrm{DS_{CM}} = \frac{161 \times A}{W - 58 \times A}$$

$$A = V_{NaOH} \times C_{NaOH}$$

where V_{NaOH} and C_{NaOH} are the volume and molarity of NaOH solution respectively, W is the weight of sample (g), and 161 and 58 are the molecular weights of glucosamine (CS skeletal unit) and carboxymethyl group respectively.

The degree of substitution of (2-hydroxy-3-trimethylammonium)propyl group (DS_{HTP}) was determined by potentiometry. The chloride ions of HTPC and *O*-CM-HTPC were potentiometrically titrated with aqueous silver nitrate using a calomel electrode as reference electrode and a silver electrode for the measurement. The method was described by Sun et al [10]. A sample (0.05 g) was dissolved in distilled water (100 ml) and the solution was titrated with 0.03N $AgNO_3$, the mV value of the solution being simultaneously recorded. The amount of aqueous $AgNO_3$ consumed was determined by second- order differential method. The degree of substitution was calculated as follows:

$$\mathrm{DS_{HTP}} = \frac{A}{A + \dfrac{W - A \times 314.5}{161}}$$

$$A = \frac{V_{AgNO_3} \times C_{AgNO_3}}{1000}$$

where V_{AgNO_3} and C_{AgNO_3} are the volume and molarity of $AgNO_3$ solution respectively, W is the weight of sample (g), and 161 and 314.5 are the molecular weights of glucosamine (CS skeletal unit) and (2-hydroxy-3-trimethylammonium)propyl chloride group respectively. The degrees of substitution (DS_{CM} and DS_{HTP}) and appearance of CS and CS derivatives are summarised in Table 1.

PHYSICAL PROPERTIES

Solubility

To estimate the solubility of a CS derivative, 100.0 mg of sample were suspended in 10.0 mL of a solvent (distilled water, 0.1N HCl or 0.1N NaOH) and the suspension was stirred at 25°C for 5 hr [17]. Then the mixture was filtered through filter paper to retain the undissolved portion, which was then washed with acetone and dried at 50˚C overnight [18]. The total weight of sample was subtracted by the weight of the insoluble portion to obtain the weight of the soluble portion. The solubility of samples was expressed as g/100 mL.

Table 1. DS_{CM}, DS_{HTP} and appearance of CS and CS derivatives

Sample	DS_{CM}[a,b]	DS_{HTP}[a,c]	Colour
CS	-	-	yellow
O-CMC	0.88 ± 0.16	-	light yellow
N,*O*-CMC	1.53 ± 0.22	-	dark yellow
HTPC	-	0.40 ± 0.24	light yellow
O-CM-HTPC	0.87 ± 0.22	0.37 ± 0.17	white

[a] average ± SD; [b] DS_{CM} for complete reaction is 1 for *O*-CMC and *O*-CM-HTPC, and 2 for *N*,*O*-CMC; [c] DS_{HTP} for complete reaction is 1 for HTPC and *O*-CM-HTPC.

As shown in Table 2, all CS derivatives showed sufficient solubility in all solvents. Interestingly, the difference in solubility of all CS derivatives in each solvent was slight, ranging between 90.0-95.5 % in 0.1N HCl, 82.6-84.2 % in distilled water and 31.6-37.6 % in 0.1N NaOH, which shows that the type and number of modification group does not seem to affect the solubility of CS derivatives.

Table 2. Solubility and apparent viscosity of CS and water-soluble CS derivatives under acidic, neutral and alkaline conditions at 25°C

Sample	0.1N HCl		Distilled water		0.1N NaOH	
	solubility (%w/v)	viscosity (mPas)	solubility (%w/v)	viscosity (mPas)	solubility (%w/v)	viscosity (mPas)
CS	72.5 ± 0.5	5.37 ±2.10	-	-	-	-
O-CMC	90.3 ± 0.5	12.10 ±3.32	82.6 ± 0.4	25.83 ±4.93	35.8 ± 1.0	18.54 ±4.05
N,*O*-CMC	92.4 ± 0.4	10.57 ±2.28	84.2 ± 0.3	22.83 ±4.44	33.8 ± 0.4	14.28 ±2.65
HTPC	90.0 ± 0.1	3.23 ±1.18	84.1 ± 0.7	3.01 ±1.18	31.6 ± 0.5	1.30 ±0.64
O-CM-HTPC	95.5 ± 0.5	4.08 ±1.31	83.9 ± 0.3	4.05 ±1.75	37.6 ± 0.6	2.64 ±1.31

Note: Values are average ± SD.

Viscosity and Rheological Property

The rheological property of a 1% (w/v) CS derivative in distilled water, 0.1N NaOH or 0.1N HCl was investigated using a Bob-and-Cup format rotational rheometer (R/S Rheometer, Brookfield Viscometer LTD., England). The measuring system was CC48 and the mode used was CSR (controlled shear rate). The measurement was done in two stages: (1) during an increase of shear rate from 0 to 1,000 s^{-1} in 3 min., and (2) during a decrease of shear rate from 1,000 to 0 s^{-1} in 3 min. All measurements were performed in triplicate at a controlled temperature of 25±1ºC. The data were analysed with Brookfield Rheo 2000 software. Apparent viscosity (Table 2), expressed in mPas, was an average of the ratio of shear stress to shear rate calculated from both the upward and downward curves of the rheogram.

Rheograms of the CS derivatives are shown in Figure 2. *O*-CMC and *N*,*O*-CMC showed higher apparent viscosity in 0.1N HCl than that of CS, and the apparent viscosity of *O*-CMC was slightly higher than that of *N*,*O*-CMC in all pH conditions studied. These observations were most likely due to the increasing hydrogen bonding and steric factor and the interplay between the two phenomena. In contrast, the cationic HTPC and *O*-CM-HTPC showed a remarkably lower viscosity than CS in 0.1N HCl and hence were less useful as a viscosity inducing agent. It should be noted that the apparent viscosity of *O*-CMC and *N*,*O*-CMC in water was significantly higher than in 0.1N HCl and 0.1N NaOH. The decrease in viscosity under alkaline condition seemed to be due to a screening effect of the counter ions (Na^+), which limits swelling at pH 9-13 [19].

O-CMC and *N,O*-CMC under acidic condition showed a Newtonian rheological property while under neutral and alkaline conditions they showed a distinct pseudoplastic property and a thixotropic hysteresis loop respectively. CS, HTPC and *O*-CM-HTPC on the other hand exhibited a Newtonian fluid property in all conditions. The shear stress of *O*-CMC in water at low shear rate (<100 sec^{-1}) was remarkably greater than that of *N*,*O*-CMC, The smaller thixotropic hysteresis area in alkaline condition compared to that in neutral condition again may be explained by the screening effect of the counter ions (Na^+) [19].

A polymer solution with pseudoplastic property and high apparent viscosity is suitable for certain pharmaceutical formulations such as suspensions and gel formulations. Thus, *O*-CMC and *N*, *O*-CMC may be preferably used as a viscosity inducing agent.

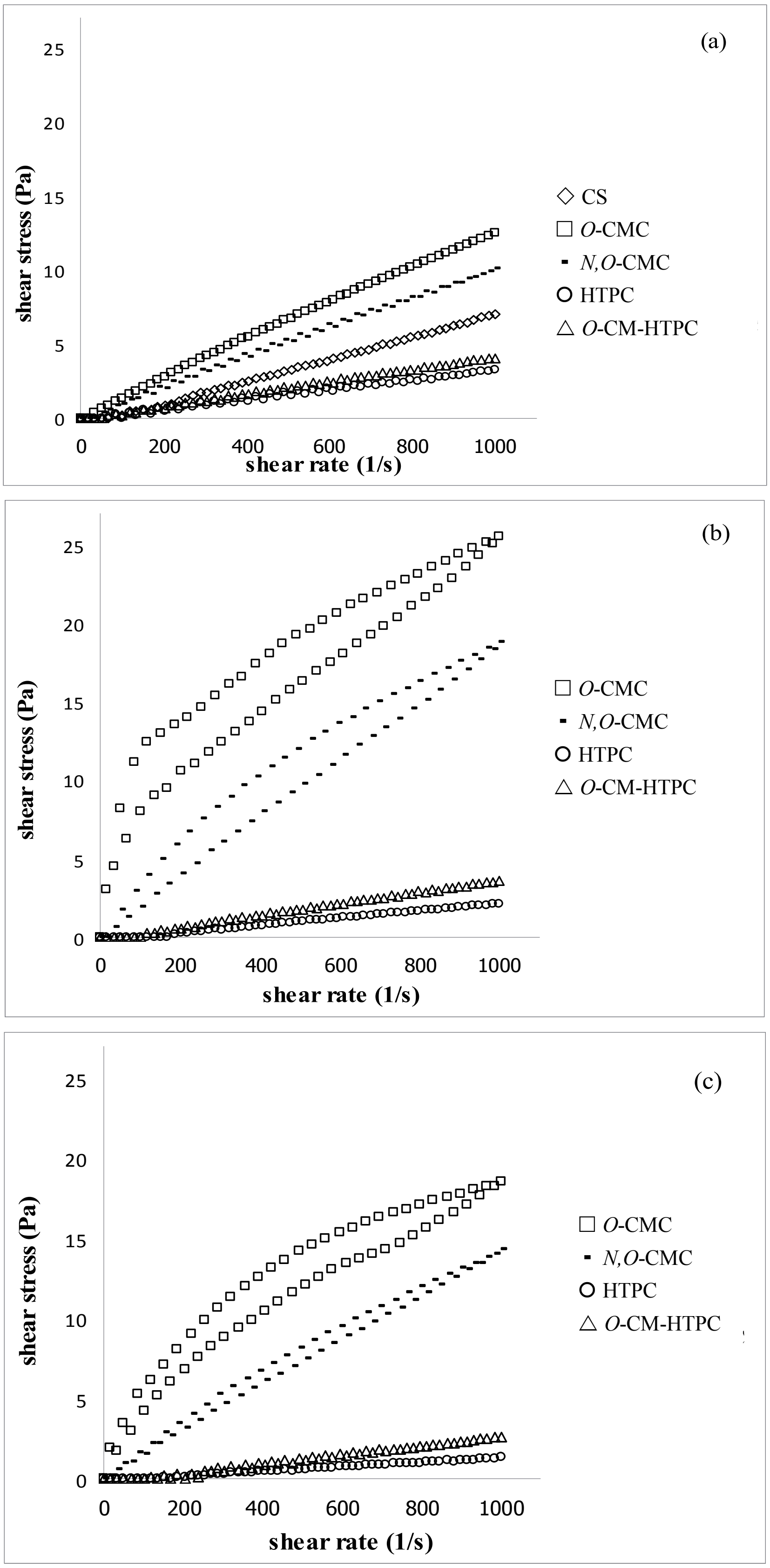


Figure 2. Rheograms of CS derivatives in: (a) 0.1N HCl; (b) distilled water; (c) 0.1N NaOH at 25°C

ACKNOWLEDGEMENTS

We thank the Department of Pharmaceutical Sciences, Faculty of Pharmacy, Chiang Mai University for providing all facilities and the Graduate School, Chiang Mai University for financial support. The Department of Chemistry, Maejo University was acknowledged for providing a potentiometer. We are grateful to Dr. Yohji Ezure for his valuable comments and proofreading of the manuscript.

REFERENCES

1. C. K. S. Pillai, W. Paul and C. P. Sharma, "Chitin and chitosan polymers: Chemistry, solubility and fiber formation", *Prog. Polym. Sci.*, **2009**, *34*, 641-678.
2. P. C. Srinivasa, M. N. Ramesh, K. R. Kumar and R. N. Tharanathan, "Properties of chitosan films prepared under different drying conditions", *J. Food Eng.*, **2004**, *63*, 79-85.
3. H. P. Brack, S. A. Tirmizi and W. M. Risen, "A spectroscopic and viscometric study of the metal ion-induced gelation of the biopolymer chitosan", *Polym.*, **1997**, *38*, 2351-2362.
4. I. Henriksen, K. L. Green, J. D. Smart, G. Smistad and J. Karlsen, "Bioadhesion of hydrated chitosans: an in vitro and in vivo study", *lnt. J. Pharm.*, **1996**, *145*, 231-240.
5. M. Rinaudo, "Chitin and chitosan: Properties and applications", *Prog. Polym. Sci.*, **2006**, *31*, 603-632.
6. G. Kerch, J. Zicans and R. M. Meri, "The effect of chitosan oligosaccharides on bread staling", *J. Cereal Sci*., **2010**, *52*, 491-495.
7. V. Cardile, G. Frasca, L. Rizza, F. Bonina, C. Puglia, A. Barge, N. Chiambretti and G. Cravotto, "Improved adhesion to mucosal cells of water-soluble chitosan tetraalkylammonium salts", *Int. J. Pharm.*, **2008**, *362*, 88-92.
8. D. de Britto and O. B. G. Assis, "A novel method for obtaining a quaternary salt of chitosan", *Carbohyd. Polym.*, **2007**, *69*, 305-310.
9. V. A. Spinelli, M. C. M. Laranjeira and V. T. Fávere, "Preparation and characterization of quaternary chitosan salt: adsorption equilibrium of chromium (VI) ion", *React Funct. Polym.*, **2004**, *61*, 347-352.
10. L. Sun, Y. Du, L. Fan, X. Chen and J. Yang, "Preparation, characterization and antimicrobial activity of quaternized carboxymethyl chitosan and application as pulp-cap", *Polym.*, **2006**, *47*, 1796-1804.
11. H. C. Ge and D. K. Luo, "Preparation of carboxymethyl chitosan in aqueous solution under microwave irradiation", *Carbohyd. Res.*, **2005**, *340*, 1351-1356.
12. A. Zhu, M. B. Chan-Park, S. Dai and L. Li, "The aggregation behavior of O-carboxymethylchitosan in dilute aqueous solution", *Coll. Surf. B: Biointerf.*, **2005**, *43*, 143-149.
13. R. Huang, E. Mendis and S. K. Kim, "Factors affecting the free radical scavenging behavior of chitosan sulfate", *Int. J. Biol. Macromol.*, **2005**, *36*, 120-127.

14. R. Xing, S. Liu, H. Yu, Z. Guo, Z. Li and P. Li, "Preparation of high-molecular weight and high-sulfate content chitosans and their potential antioxidant activity in vitro", *Carbohyd. Polym.*, **2005**, *61*, 148-154.
15. E. R. Hayes, "N,O-carboxymethyl chitosan and preparative method therefore", U.S. Patent 4,619,995, Oct 28, **1986.**
16. X. Liang, H. Wang, H. Tian, H. Luo and J. Chang, "Synthesis, structure and properties of novel quaternized carboxymethyl chitosan with drug loading capacity", *Acta Phys. Chim. Sin.*, **2008**, *24*, 223-229.
17. Y. C. Chung, C. L. Kuo and C. C. Chen, "Preparation and important functional properties of water-soluble chitosan produced through Maillard reaction", *Bioresour. Technol.*, **2005**, *96*, 1473-1482.
18. Y. Xie, X. Liu and Q. Chen, "Synthesis and characterization of water-soluble chitosan derivate and its antibacterial activity", *Carbohyd. Polym.*, **2007**, *69*, 142-147.
19. M. Sadeghi, "Synthesis and swelling behaviors of graft copolymer based on chitosan-g-poly (AA-co-HEMA) ", *Int. J. Chem. Eng. Appl.*, **2010**, 1, 354-358.

Maejo Int. J. Sci. Technol. **2012**, *6*(02), 323-333

Maejo International
Journal of Science and Technology

ISSN 1905-7873
Available online at www.mijst.mju.ac.th

Technical Note

Development of a rehabilitation apparatus to actuate upper extremity passive motion

Chung-shun Hung [1], Ching-hua Chiu [2,*], Yi-hung Ho [1] and Wei-lin Cheng [1]

[1] Department of Bio-Industrial Mechatronics Engineering, The College of Agriculture and Natural Resources, National Chung Hsing University, 250 KuoKuang Road, Taichung, Taiwan 402, the R.O.C.

[2] Graduate Institute of Sports and Health Management, The College of Social Sciences and Management, National Chung Hsing University, 250 KuoKuang Road, Taichung, Taiwan 402, the R.O.C.

* Corresponding author, e-mail: chungoodman@yahoo.com.tw

Received: 26 June 2011 / Accepted: 28 August 2012 / Published: 29 August 2012

Abstract: An apparatus that can induce upper extremity passive motion in stroke patients was developed with the goal of providing rehabilitation for these individuals. The rehabilitation device consists of a robotic arm controlled by a computer interface and programmed to effect passive extension and flexion of the patient's elbow and fingers. The load imposed on the upper limb and fingers was analysed. A high-speed video camera captured the trajectory of the subject's arms and a kinetic model based on the arm structure was employed to analyse the trajectory. Results showed that the range of motion (ROM) of the subjects' elbows was between 77-158° (about 80° ROM) and the average range of movement from extension to flexion of the fingers was approximately 3 cm. The average loading moment on the shoulder and elbow was below 12 N-m. By varying the angular position and angular velocity of the robotic arm, the programme reproduced the motions of different arm functions. Thus, the present study shows that the apparatus provides safe and effective simultaneous rehabilitation of both the elbow and fingers, and that the use of this device may benefit patients undergoing rehabilitation in a clinical setting.

Keywords: rehabilitation apparatus, post-stroke patient, upper extremity passive motion

---

INTRODUCTION

Stroke occurs when circulation of the blood is suddenly blocked, leading to oxygen deficiency in the brain. The effect of stroke is closely related to the distribution of the patients' brain vessels, the extent of concomitant disease processes, and the age of the patients at the onset of stroke. There are two types of stroke: hemorrhagic and nonhemorrhagic [1]. Stroke sufferers may experience partial paralysis of some limbs and in serious cases may totally lose function in affected body parts. One possible way to recover the original function in an affected limb is through targeted exercise with the assistance of an orthosis.

Jorgenson et al. [2] reported that two-thirds of stroke patients experienced paralysis of their upper limbs and were unable to move them freely. Even after 6 months of rehabilitation, more than 50% of those patients still felt no strength in their upper limbs and were incapable of moving them autonomously [3]. In a typical stroke case, the patient cannot maintain the flexion and extension of the elbow, wrist and fingers. Consequently, much research has focused on the use of gravity support to help patients effectively stretch their fingers and relax the finger flexor.

Housman et al. [4] observed that when a physiotherapist offered gravity assistance with a nonrobotic (despite the name) therapy Wilmington robotic exoskeleton (T-WREX), the patient's arm became more flexible, paralysis was relieved, and the arm's passive motion range significantly increased. In a study by Seo et al. [5], the grip strength of stroke patients trained with proper arm support was greatly enhanced. Meijer et al. [6] designed a device, the Handmaster orthosis, which was intended to help patients flex and extend their wrists. Their findings indicated that patients who used the device for only a short period experienced a remarkable increase in the passive range of motion (ROM). Similar results were also reported by Alon et al. [7], who found that the Handmaster system could improve the grip strength and the active ROM of patients' fingers.

After observing patients' conditions, doctors and physiotherapists tend to have patients engage in rehabilitation to improve muscle strength and ROM of joints. However, objective monitoring of rehabilitation progress is difficult, especially when the number of patients under assessment is very large. Therefore, robot-assisted rehabilitation has become increasingly popular, and research and development efforts in pursuit of such devices have rapidly expanded. For example, Mehrholz et al. [8] reviewed 11 trials in which electromechanical and robot-assisted arm training devices were used to improve arm function and activities of daily life. Their findings revealed that although the arm motor function and strength improved, significant improvement in the activities of daily life was lacking. Robots designed to assist stroke patients in recovering their arm muscle strength and autonomous motion have focused on activities that improve immobile strength and endurance, and results of clinical testing and follow-up diagnoses by physicians have confirmed previous conclusions that such rehabilitative efforts produced their intended effects [9-10].

From our literature review, numerous methods aimed at rehabilitating the upper extremities of patients have already been investigated. These approaches to rehabilitation concentrated on recovering patients' mobility, especially by improving the muscle strength and flexibility of the arms and fingers. However, most of the rehabilitation devices described were not versatile because they were designed to act on a single joint or body part. In view of this, the present study attempts to accomplish the following two tasks: (1) development of a convenient rehabilitation apparatus for the upper extremities that can produce passive flexion and extension of patients' elbows and fingers; and (2) examination of the load imposed on the patients' arms through analysis of the passive flexion and extension movements of the elbows and fingers.

METHODS

Abbreviations and Definitions

$\boldsymbol{\tau}_i$ = generalised force
q_i = generalised coordinate
$\dot{\boldsymbol{q}}_i$ = generalised velocity
$\ddot{\boldsymbol{q}}_i$ = generalised acceleration
${}^{i-1}E_i$ = homogeneous transformation matrix of ith coordinate frame relative to i-1th coordinate frame (see Appendix A)
${}^{0}E_j$ = coordinate transformation matrix from 0 coordinate frame to jth coordinate frame, and the equation could be written as ${}^{0}A_j = {}^{0}A_1{}^{1}A_2{}^{2}A_3 \ldots {}^{j-1}A_j$
$\boldsymbol{U}_{ij}$ is defined as $({}^{0}E_i/(q_j$ $(i,j = 1,2,\ldots,5)$
$\boldsymbol{U}_{ijk}$ is defined as $({}_{0}\boldsymbol{U}_{ij}/(q_k$ $(i,j,k = 1,2,\ldots,5)$
$\boldsymbol{J}_i$ = pseudo-inertia matrix 29–32
$G = [0,0,-|g|,0]$, $g = 9.8062$ ms^{-2}
$\boldsymbol{m}_j$ = mass of jth link
$\boldsymbol{r}_j = (\bar{x}_j, \bar{y}_j, \bar{z}_j, 1)^T$, position of centre of mass for jth link

Kinematic Model of Upper Extremity

The kinematic model of the human upper extremity used in this study is composed of five degrees of freedom (q_i; $i = 1, 2, \ldots, 5$), based on the structure of the arm joints. The shoulder joint (q_3) comprises three degrees of freedom (Figure 1). Two degrees of freedom describe the translation relative to the X-axis (q_1) and Y-axis (q_2), and the third is associated with the rotation of the shoulder joint. The fourth degree of freedom reflects the rotation of the elbow joint (q_4) and the fifth is associated with the rotation of the wrist joint (q_5). The symbol θ represents the joint angle of the elbow. The segment lengths of the upper arm, forearm and palm are represented by $\boldsymbol{l}_1$, $\boldsymbol{l}_2$ and $\boldsymbol{l}_3$ respectively. In Figure 1, $\boldsymbol{p}_i$ represents the translation from the origin of the coordinate frame of the ith segment to the coordinate frame of the i-1th link; $\boldsymbol{p}_i = [x_i, y_i, z_i]^T$, and r_i represents the position vectors of the centre of mass of segment i.

The five degrees of freedom serve as variables in this upper extremity model (Table 1). Each degree of freedom is assumed to be a coordinate frame. Consequently, according to the Lagrange-Euler equations of motion [11-16], this model can be represented as:

$$\tau_i = \sum_{j=i}^{n} \sum_{k=1}^{j} \mathrm{Trace}(\boldsymbol{U}_{jk}\boldsymbol{J}_j\boldsymbol{U}^T{}_{ji})\, \ddot{q}_k + \sum_{j=i}^{n} \sum_{k=i}^{j} \sum_{m=1}^{j} \mathrm{Trace}(\boldsymbol{U}_{jkm}\boldsymbol{J}_j\boldsymbol{U}^T{}_{ji})\, \dot{q}_k\, \dot{q}_m$$
$$- \sum_{j=i}^{n} (\boldsymbol{m}_j\boldsymbol{G}\boldsymbol{U}_{ji}\, r_j) \quad i,j,k = 1,2,3,\ldots,5. \qquad (1)$$

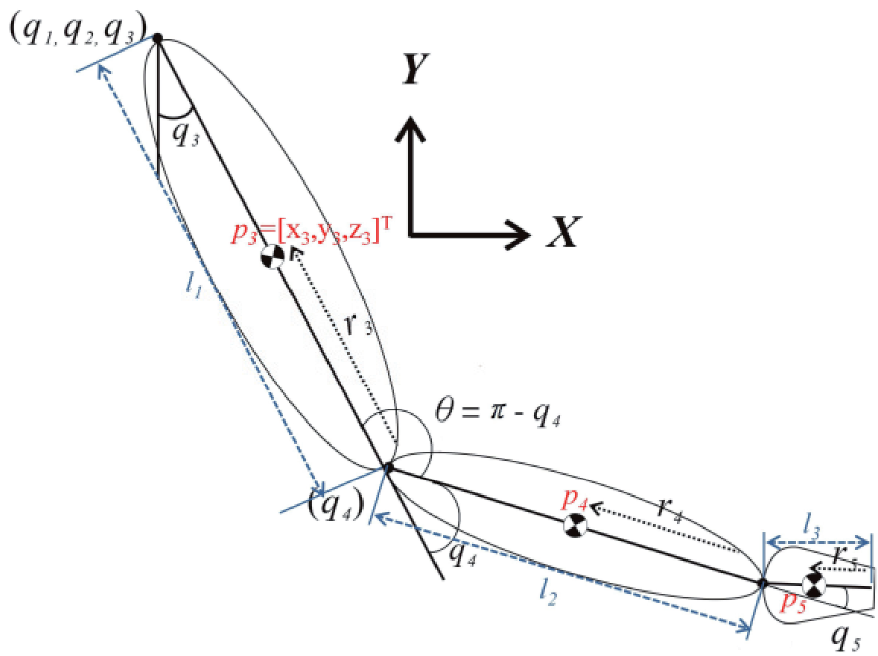


Figure 1. Upper extremity model

Table 1. Homogeneous transformation matrices and parameters for the arm model (see Appendix A)

Variable	${}^{i-1}E_i$ type	x_i	y_i	z_i	${}^0E_i$
q_1	${}^0E_1 = E_1^{(x)}$	0	0	0	${}^0E_1 = E_1^{(x)}$
q_2	${}^1E_2 = E_2^{(y)}$	0	0	0	${}^0E_2 = E_1^{(x)}\, E_2^{(y)}$
q_3	${}^2E_3 = E_3^{(R)}$	0	l_1	0	${}^0E_3 = E_1^{(x)}\, E_2^{(y)}\, E_3^{(R)}$
q_4	${}^3E_4 = E_4^{(R)}$	0	l_2	0	${}^0E_4 = E_1^{(x)}\, E_2^{(y)}\, E_3^{(R)}\, E_4^{(R)}$
q_5	${}^4E_5 = E_5^{(R)}$	0	l_3	0	${}^0E_5 = E_1^{(x)}\, E_2^{(y)}\, E_3^{(R)}\, E_4^{(R)}\, E_5^{(R)}$

The segment parameters required by the dynamic system include length, mass and inertia tensor. Zatsiorsky and Seluyanov [17] adopted a gamma-ray scanning technique in their study on human segment parameters. Compared with other studies, their data were more precise and complete. Therefore, the segment parameters used in the dynamic system of this study are based on the data established by Zatsiorsky and Seluyanov [17].

Subjects and Rehabilitation Apparatus

Fifteen stroke patients from China Medical University Hospital (age: 61.7 ± 11.5 years; height: 1.62 ± 0.08 m; weight: 64.6 ± 10.0 kg) participated in the study. All of the participants provided written consent to the experimental protocol approved by the institutional review board.

The rehabilitation apparatus used in the present study is designed for rehabilitation of the elbow and fingers. The mechanical components related to elbow rehabilitation include the axis of rotation, link bar, starting device and elbow bearer (Figure 2). To ensure the safety of the subject when using the apparatus, a limit-switch sensor is attached to the axis of rotation to prevent elbow injury due to over-bending.

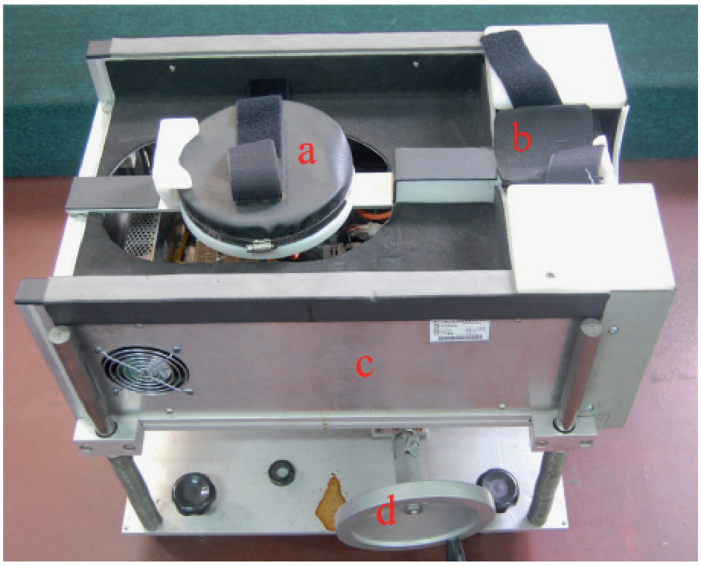


Figure 2. Rehabilitation apparatus: (a) filling solenoid, (b) elbow bearer, (c) axis of rotation, link bar and starting device in box, and (d) lifting mechanism

The components of the apparatus for elbow rehabilitation were carefully designed with comfort and safety in mind. Patients using this apparatus begin by resting the elbow on a plate (Figure 2, speed range = 0.1-0.8 ± 0.01 rad/s, range of joint = 0-90 ± 0.1°, and maximum joint moment = 45 ± 0.1 N-m for this apparatus). The height of the plate can be adjusted for maximal comfort.

The major mechanical components involved in fingers rehabilitation comprise a gas pocket and filling solenoid (Figure 2(a)). Patients or users wear velcro gloves and their hand and fingers are attached to the filling solenoid with velcro. The filling solenoid is connected to a gas pocket in order to inflate it at intervals, allowing repeated flexion of the patient's fingers.

The control interface of this apparatus was developed in Borland C^{++} programming language and the operating procedure for this apparatus is shown in Figure 3. The transmission of communication signals for motor control is via a universal serial bus (USB) interface. After starting the apparatus, the user enters the control parameters into the computer. These parameters comprise extension frequency, flexion frequency and joint angles. The input parameters function as the preconditions for controlling the elbow's initial angular position and angular velocity, and also serve as parameters for the motion procedures. The user then starts the motor to inflate the gas solenoid, compelling the fingers to extend and flex.

In these experiments, a high-speed video camera (120 Hz) was positioned at the side of the subject to capture upper extremity motions in two dimensions (for example, Figure 4). Five markers were placed at the following anatomical positions: right third phalanx, right styloid process of the radius, right lateral epicondyle of the humerus, right acromion joint, and the lateral midpoint of trunk. The markers' locations in the captured video were processed digitally. A programme developed in Borland C++ language was used to compute the angular position, angular velocity and passive loading moment of the elbow and shoulder joints.

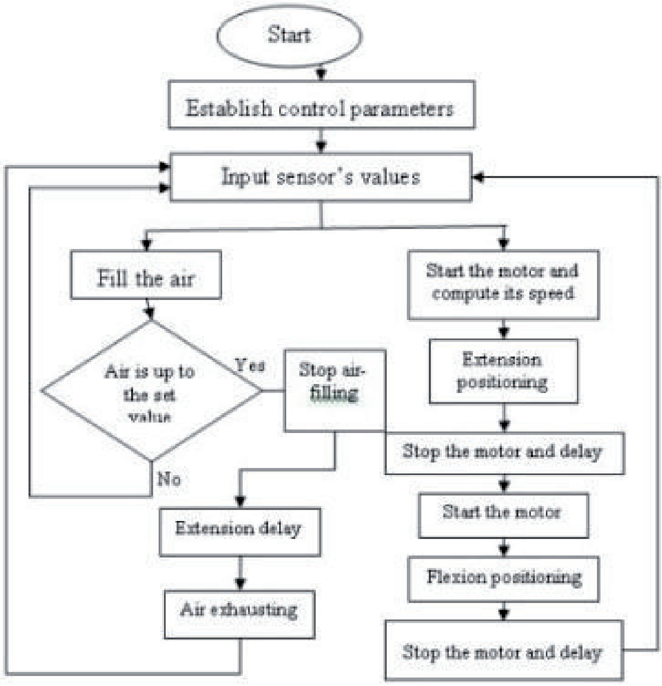
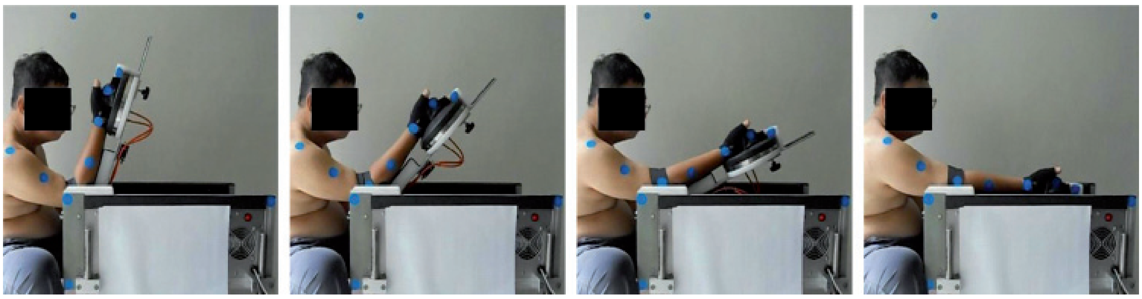


Figure 3. Procedure for using the apparatus



(a) (b) (c) (d)

Figure 4. Consecutive rehabilitation motions: (a) initial state, in which control parameters are input into computer; apparatus is then started; (b, c) extension of elbow joint; (d) extension of elbow to maximal angle. As elbow is extended, gas pocket is inflated to compel fingers to extend until maximal elbow extension angle is reached. The procedure is then reversed from (d) to (c), (b) and finally (a). The motion can be repeated to achieve continuous rehabilitation.

RESULTS AND DISCUSSION

When the subject's elbow was extended by the robotic arm, the gas pocket was inflated to compel passive plantarflexion of the subject's fingers. The markers' positions were captured by the video motion system and used to compute the height of gas pocket during filling and exhausting (Figure 5). The angular position, angular velocity and passive loading moment of the elbow and shoulder joints were calculated using programmes. A collected data sample from one patient is shown in Figure 6. The joint motions of shoulder and elbow for rehabilitation were observed and evaluated. Then, the experimental results of all subjects were calculated.

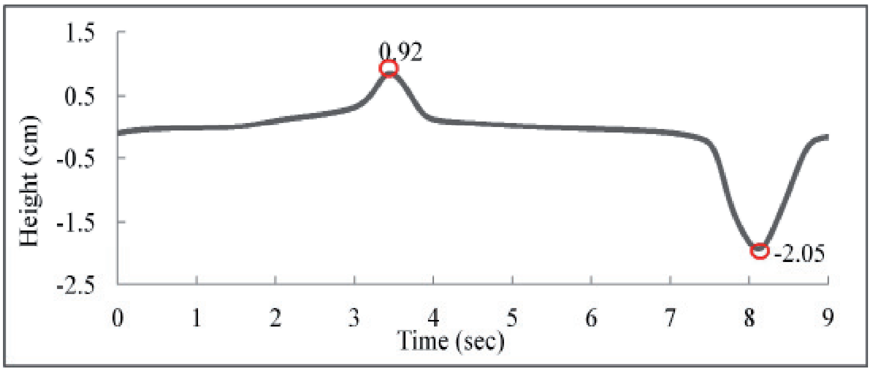


Figure 5. Average height of gas pocket during filling and exhausting (maximum: 0.92 ± 0.24 cm within 3.4 ± 0.11 s; minimum: –2.05 ± 0.46 cm within 8.2 ± 0.1 s).

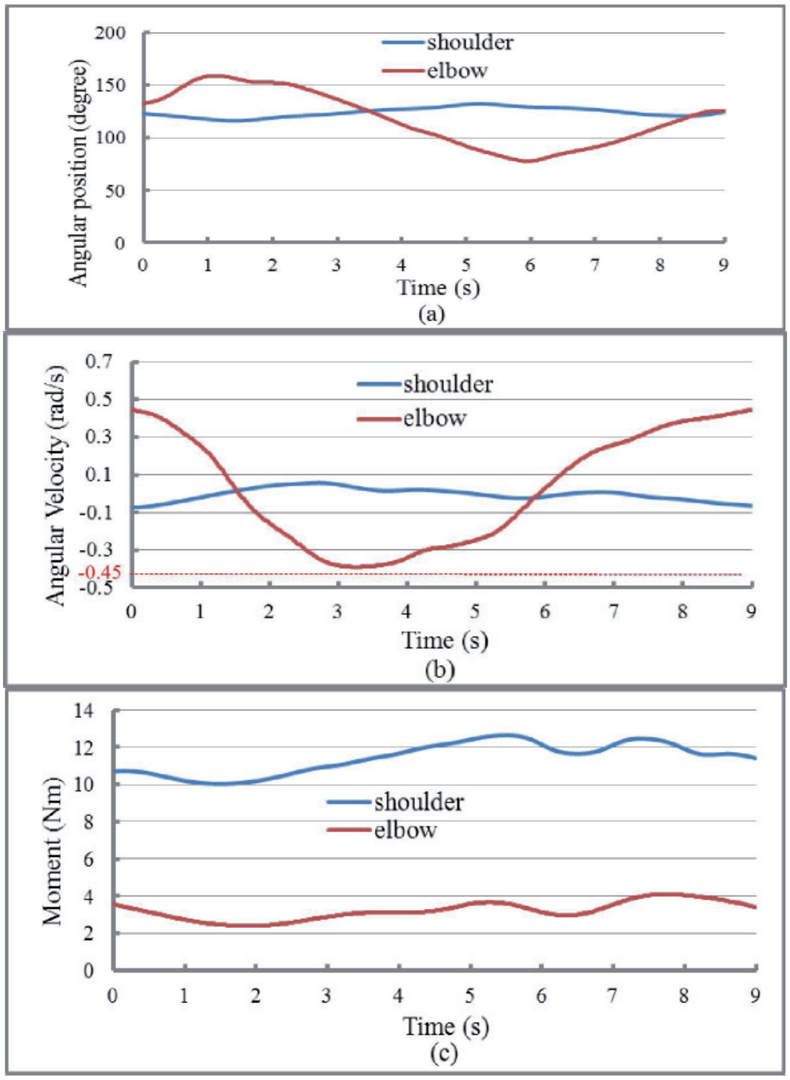


Figure 6. Sample results from one subject: (a) shoulder's angular position (q_3) and elbow's angular position (q_4); (b) angular velocities ($\dot{q}_3$, $\dot{q}_4$) of shoulder and elbow ; (c) passive moment of shoulder and elbow joints

According to the study by Wade et al. [18], stroke patients usually experience severe arm paralysis at the initial rehabilitation stage. To regain the level of function prior to the stroke, patients normally require at least 6 months of rehabilitation, during which time the arm needs to be constantly mobilised [19]. In the process of flexing and extending the shoulder and elbow joints provided by our apparatus (for example, Figure 6(a)), little variation was observed in the shoulder flexion angle whereas the elbow underwent more variation in flexion angle. The apparatus mainly provided more rehabilitation activity in the elbow than in the shoulder. The elbow activity could then provide muscle rehabilitation of the upper arm and forearm. In the present study the subjects' range of elbow motion, i.e. the average flexion angle (θ) for the elbow, was from $77.6 \pm 8.1°$ to $158.6 \pm 14.6°$ (about 80° ROM), which is within the range of ordinary elbow motions [14] and thus unlikely to cause injury.

The programmed shoulder and elbow angular velocities during extending were both below -0.45 rad/s (negative angular velocity) (as in Figure 6(b)), which are slower than the velocities during normal activity [14]. For rehabilitating stroke patients, sudden and dramatic motions are not appropriate. Therefore, the device should be stabilised when varying the ROM of the robotic arm [8].

We found that the profiles of the passive moment applied to the elbow and shoulder joints were similar (Figure 6(c)) because both joints were compelled to move by the robotic arm. The average moment of the subjects' shoulders was 11.2 ± 2.11 N-m (below 12 N-m), which is almost equivalent to the load induced by ordinary arm activity in daily life [12]. Because increased joint moment is helpful for neuromuscular strengthening, increasing joint moment is one of the rehabilitation device's functions. The apparatus compels the arm to perform motions in a repeated and fixed pattern, effectively stimulating muscles and cutaneous receptors, thereby allowing patients to gain better neuromuscular control. With the aid of this apparatus, therefore, the patients seemed to be able to increase their joint moment and recover their muscle strength.

The apparatus also provides rehabilitation motions for fingers (Figure 2(a)) to promote grip strength and increase their range of movement [7]. Periodic inflation of the gas pocket facilitates passive extension and flexion of the fingers (Figure 5). When the robotic arm returns to its initial position, the fingers are returned to passive dorsiflexion. In the present study, the range of extension and flexion of the fingers is controlled by the height of the gas pocket, which changes from 0.92 ± 0.24 cm to -2.05 ± 0.46 cm, the full average rehabilitation range from extension to flexion being approximately 3 cm. The range from extension to flexion is used to assess the motion activity of fingers and the progress of muscle rehabilitation.

There are some advantages to using a mechanical rehabilitation apparatus in place of manual passive motion by a therapist. In addition to the operational costs being reduced, rehabilitative motions will be more reproducible and technical errors will be less likely to occur [20]. The rehabilitation apparatus designed in the present study is intended for use with stroke patients. With this apparatus, motions of patients' upper extremities can be computer-controlled. Such controlled activities may help patients recover arm strength and active ROM by inducing motions of specified magnitude, duration and time [9-10]. The USB connection was adopted because of its 'plug and play' characteristics, which improve convenience and reduce operator errors. Our apparatus and its control programme allow users to easily monitor rehabilitation progress and adjust the control parameters and operating regime accordingly on the computer.

Most rehabilitation devices are intended to operate on a single body part. For example, Dobbe et al. [21] invented a finger rehabilitation device employing a constant-force spring motor. Our apparatus offers dual function, rehabilitating the elbow and fingers simultaneously. Efficiency is enhanced because less rehabilitation time is required. It can also offer different motion programmes with variation in ROM and angular velocity without risking injury to the patients. These motion programmes and operational functions can be applied in clinical practice because they are safe and convenient to use.

CONCLUSIONS

The apparatus described in this report is safe and helpful for rehabilitating patients. With this apparatus, rehabilitation courses can be more diverse and thus more appealing. Future research should explore how the factors affecting the speed and range of the mechanical device optimise rehabilitation of the upper extremities. Forthcoming efforts should also include promotion of the device's efficacy and comfort when used in clinical practice, leading to enhanced rehabilitative effects.

ACKNOWLEDGEMENTS

The help of the fifteen stroke patients and the physiotherapists from China Medical University Hospital are truly appreciated.

REFERENCES

1. H. P. Adams Jr., B. H. Bendixen, L. J. Kappelle, J. Biller, B. B. Love, D. L. Gordon and E. E. Marsh III, "Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment", *Stroke*, **1993**, *24*, 35-41.
2. H. S. Jorgensen, H. Nakayama, H. O. Raaschou and T. S. Olsen, "Stroke. Neurologic and functional recovery: the Copenhagen stroke study", *Phys. Med. Rehabil. Clin. N. Am.*, **1999**, *10*, 887-906.
3. G. Kwakkel, B. J. Kollen, J. van der Grond and A. J. H. Prevo, "Probability of regaining dexterity in the flaccid upper limb: Impact of severity of paresis and time since onset in acute stroke", *Stroke*, **2003**, *34*, 2181-2186.
4. S. J. Housman, K. M. Scott and D. J. Reinkensmeyer, "A randomized controlled trial of gravity-supported, computer-enhanced arm exercise for individuals with severe hemiparesis", *Neurorehabil. Neural Repair*, **2009**, *23*, 505-514.
5. N. J. Seo, W. Z. Rymer and D. G. Kamper, "Delays in grip initiation and termination in persons with stroke: effects of arm support and active muscle stretch exercise", *J. Neurophysiol.*, **2009**, *101*, 3108-3115.

6. J. W. G. Meijer, G. E. Voerman, K. M. L. W. Santegoets and A. C. H. Geurts, "Short-term effects and long-term use of a hybrid orthosis for neuromuscular electrical stimulation of the upper extremity in patients after chronic stroke", *J. Rehabil. Med.*, **2009**, *41*, 157-161.
7. G. Alon, K. McBride and H. Ring, "Improving selected hand functions using a noninvasive neuroprosthesis in persons with chronic stroke", *J. Stroke Cerebrovasc. Dis.*, **2002**, *11*, 99-106.
8. J. Mehrholz, T. Platz, J. Kugler and M. Pohl, "Electromechanical and robot-assisted arm training for improving arm function and activities of daily living after stroke", *Stroke*, **2009**, *40*, 392-393.
9. T. G. Sugar, J. He, E. J. Koeneman, J. B. Koeneman, R. Herman, H. Huang, R. S. Schultz, D. E. Herring, J. Wanberg, S. Balasubramanian, P. Swenson and J. A. Ward, "Design and control of RUPERT: a device for robotic upper extremity repetitive therapy", *IEEE Trans. Neural Syst. Rehabil. Eng.*, **2007**, *15*, 336-346.
10. P. Lam, D. Hebert, J. Boger, H. Lacheray, D. Gardner, J. Apkarian and A. Mihailidis, "A haptic-robotic platform for upper-limb reaching stroke therapy: preliminary design and evaluation results", *J. Neuroeng. Rehabil.*, **2008**, *5*, 1-13.
11. C. H. Chiu, "The preliminary study of optimal planning for front chin-ups", *J. Med.Biol. Eng.*, **2005**, *25*, 129-135.
12. C. H. Chiu, W. T. Hsul, M. H. Hsu, S. L. Chuang, J. C. Wang and T. Y. Hsu, "A numerical estimation of energy expenditure for wheelchair users", *J. Med. Biol. Eng.*, **2007**, *27*, 79-86.
13. C. H. Chiu, "Computer algorithm of optimal planning for the support of a single leg", *Int. J. Sport Exerc. Sci.*, **2009**, *1*, 1-8.
14. C. H. Chiu, J. C. Wang and C. F. Chen, "Minimum energy expenditure of arm and leg motions", *Int. J. Sport Exerc. Sci.*, **2009**, *1*, 33-40.
15. C. H. Chiu, "Computer algorithm of optimal planning for single-hand suspension motion", *Int. J. Sport Exerc. Sci.,* **2009**, *1*, 59-67.
16. K. S. Fu, R. C. Gonzalez and C. S. Lee, "Robotics: Control, Sensing, Vision, and Intelligence", McGraw-Hill, New York, **1987**, Ch.2-6.
17. V. Zatsiorsky and V. Seluyanov, "The mass and inertia characteristics of the main segments of the human body", in "Biomechanics VIII-B" (Ed. H. Matsui and K. Kobayashi), Human Kinetics Publishers, Chamapaign, IL., **1983**, pp.1152-1159.
18. D. T. Wade, R. Langton-Hewer, V. A. Wood, C. E. Skilbeck and H. M. Ismail, "The hemiplegic arm after stroke: measurement and recovery", *J. Neurol. Neurosurg. Psychiatry*, **1983**, *46*, 521-524.
19. F. N. Shelton and M. J. Reding, "Effect of lesion location on upper limb motor recovery after stroke", *Stroke*, **2001**, *32*, 107-112.
20. L. Oujamaa, I. Relave, J. Froger, D. Mottet and J. Y. Pelissier, "Rehabilitation of arm function after stroke. Literature review", *Ann. Phys. Rehabil. Med.*, **2009**, *52*, 269-293.
21. J. G. G. Dobbe, N. E. van Trommel and M. J. P. F. Ritt, "Patient compliance with a rehabilitation program after flexor tendon repair in zone II of the hand", *J. Hand Ther.*, **2002**, *15*, 16-21.

22. C. H. Chiu, "The Study of the dynamic modeling for the support phase in running", *PhD Thesis*, **1999**, National Taiwan Normal University, Taiwan.

APPENDIX A

Chiu [22] defined seven types (*CH-7T*) of homogeneous transformation matrices [11-15]. This study adopted three of the seven types to design the LE equations. For example, in the basic homogeneous rotation matrix (Table 1) using equations (2–4), the symbol q_i represents a generalised coordinate as a joint variable associated with the ith link. The translation from the origin of the ith link coordinate frame relative to the i-1th link coordinate frame is represented by p_i, where $p_i = [\, x_i, y_i, 0\,]^T$.

$$E_i^{(x)} = \begin{bmatrix} 1 & 0 & 0 & q_i \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (2)$$

$$E_i^{(y)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & q_i \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (3)$$

$$E_i^{(R)} = \begin{bmatrix} \cos q_i & -\sin q_i & 0 & 0 \\ \sin q_i & \cos q_i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & x_i \\ 0 & 1 & 0 & y_i \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (4)$$